

GRGETNTGADKSKSLADGMSLLGAILVHHGIEVARQVILRFGTLLDAAPTLAGL
 DWKTSLOBLTARGMGVPSYVYVSTGPDHDKFTAVVVMVDTEVSGIGHSKKEAEQK
 AASAAKALDVLGGVGTSTV
 complement(1980. .2177)
 /gene="rnc"
 /note="pfam match to entry PF00035 dsm, Double-stranded
 RNA binding motif, score 84.20, E-value 2.7e-21"
 complement(2259. .2540)
 /gene="rnc"
 /note="pfam match to entry PF00636 Ribonuclease_3, RNase3
 domain, score 155.60, E-value 8.2e-43"
 complement(2514. .2540)
 /gene="rnc"
 /note="PS00517 Ribonuclease III family signature"
 complement(2657. .3310)
 /gene="ML1660"
 complement(2657. .3310)
 /gene="ML1660"
 /note="Similar to several proteins of undefined function
 e.g. Mycobacterium tuberculosis RV2926C TR:YT26_MYCTU."
 /codon_start=1
 /product="conserved hypothetical protein"
 /protein_id="CAC30613.1"
 /db_xref="GI:13093431"
 /db_xref="SWISS-PROT:O69468"
 /translation="MTVALRLDCRISTMTQYGVTTQRLHLSIPVFDITPLGRRP
 GAITALQKTVSLRIEGLVIEWGAFLINDLREVSESDVLVAGTVTPTVSECVR
 CLTAVHGVQVTLNQLFAYPSATKTEEDAVGHVVDGTDLEQSIIDAVGLELPPA
 PMCRSDCPGLCAGCGTSLVVEPQPHDRIDPWNAKLDMPLADVPQPTSETDGRSEW"
 complement(3344. .4087)
 /gene="ML1661"
 complement(3344. .4087)
 /gene="ML1661"
 /note="Similar to several proteins of undefined function
 e.g. Mycobacterium tuberculosis RV2927C TR:YT27_MYCTU."
 /codon_start=1
 /product="conserved hypothetical protein"
 /protein_id="CAC30614.1"
 /db_xref="GI:13093432"
 /db_xref="SWISS-PROT:Q9CB56"
 /translation="MAVYVFEALDELGAIVERAGVPMTAGCVPRGDVLELIDDIK
 DAIPGELDDQDLVDARDMLNDKADHSMVSATSESLISHARAEADRLISAK
 SQVDMASERQHSERVLGDARSHRIATVAKREVEASLNRAQSCDRDLENGNISY
 EFLNGTILSVGRGHQRTAAGTHDYVTR"
 complement(4090. .4095)
 /gene="ML1662"
 /pseudo
 4317. .4897
 /gene="ML1662"
 /note="Possible pseudogene of M. tuberculosis orthologue
 RV2963 (Best blastx score 136)"
 /pseudo
 /codon_start=1
 /transl_table=11
 /product="integral membrane protein (pseudogene)"
 4905. .6083
 /note="1179 bp 100% identical to region 545582. .544404"
 complement(6182. .6655)
 /gene="coad"
 /note="synonym: ML1663"
 complement(6182. .6655)
 /gene="coad"
 /note="Highly similar to many phosphopantetheine
 adenylyltransferases (EC 2.7.7.3) involved in coenzyme A
 biosynthesis, including: Escherichia coli SW:COAD_ECOLI
 (P23875) (159 aa); Fasta score E(): 5.9e-25, 43.5%
 identity in 154 aa overlap and Mycobacterium tuberculosis
 RV2965C SW:COAD_MYCTU (Q50452) (161 aa); Fasta score E():
 0. 85.6% identity in 153 aa overlap. Contains Pfam match

to entry PF01467 Cytidylyltransf, Cytidylyltransferase."
 /codon_start=1
 /transl_table=11
 /product="probable phosphopantetheine adenylyltransferase"
 /protein_id="CAC30616.1"
 /db_xref="GI:13093433"
 /db_xref="GOA:O69466"
 /db_xref="SWISS-PROT:O69466"
 /translation="WVCPGSPFPTLGHIDVFERRAAQDFEVVAILNPVKGMFPL
 DERAMINSTHLENLREVEAGEVLVALVRGKMTAIVKGLRTGVDFEYELQMAQMN
 KHIAGVDTFVFATAPRYSFVSSSLVKVEAMLGDSVSELPSVNRFRKNGSTS"
 complement(6257. .6652)
 /gene="coad"
 /note="Pfam match to entry PF01467 Cytidylyltransf,
 Cytidylyltransferase, score 157.90, E-value 1.7e-43"
 complement(6687. .7160)
 /gene="ML1664"
 complement(6687. .7160)
 /gene="ML1664"
 /note="Similar to several proteins of undefined function
 including: Neisseria meningitidis possible DNA methylase
 TR:CAE83466 (EMBL:AL162752) (189 aa); Fasta score E():
 3.6e-07, 39.6% identity in 96 aa overlap and Mycobacterium
 tuberculosis RV2966C TR:P95128 (EMBL:Z83018) (188 aa);
 Fasta score E(): 0. 72.9% identity in 155 aa overlap.
 Contains PS00092 N-6 Adenine-specific DNA methylases
 signature."
 /codon_start=1
 Query Match 92.0%; Score 18.4; DB 1; Length 348950;
 Best Local Similarity 95.0%; Pred. No. 4.4e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GGCCGCGACCACTCGACTT 20
 |||||
 DB 1854 GGCCGCGACCACTCTACTT 1873
 |||||
 RESULT 14
 SNO224512/c
 LOCUS SNO224512 19632 bp DNA linear BCT 20-OCT-1999
 DEFINITION Streptomyces nogalater nogalamycin biosynthesis genes.
 ACCESSION AJ224512 248262
 VERSION AJ224512.2 GI:6093257
 KEYWORDS 2,3-dehydratase; 3,5-epimerase; activator; aklanonic acid;
 aronates; dTDP-4-keto-6-deoxyhexose reductase; epimerase;
 glycosyltransferase; methylase; N-methylase; nogalamycin resistance;
 O-methylase; oxygenase; polyketide; polyketide ketoreductase; snoa1
 gene; Snoa1 protein; snoa2 gene; Snoa2 protein; snoa3 gene; Snoa3
 protein; snoaB gene; SnoaB protein; snoaC gene; SnoaC protein;
 snoad gene; Snoad protein; snoas gene; Snoas protein; snogf gene;
 snogf protein; snogg gene; snogg protein; snogh gene; snogh
 protein; snogX gene; snogX protein; snogY gene; snogY protein;
 snogZ gene; snogZ protein; snop gene; snop protein; snorA gene;
 snorA protein; snoro gene; snoro protein.
 Streptomyces nogalater
 Streptomyces nogalater
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.
 1
 Ylihkonko,K., Tuikkanen,J., Jussila,S., Cong,L. and Mantsala,P.
 A gene cluster involved in nogalamycin biosynthesis from
 Streptomyces nogalater: sequence analysis and complementation of
 early-block mutations in the anthracycline pathway
 Mol. Gen. Genet. 251 (2), 113-120 (1996)
 96242142
 8668120
 2
 Ylihkonko,K., Hakala,J., Kunnari,T. and Mantsala,P.
 Production of hybrid anthracycline antibiotics by heterologous
 expression of Streptomyces nogalater nogalamycin biosynthesis genes
 Microbiology (Reading, Engl.) 142 (Pt 8), 1965-1972 (1996)
 96349102
 SOURCE
 ORGANISM
 Streptomyces nogalater
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Streptomycineae; Streptomycetaceae; Streptomyces.
 1
 Ylihkonko,K., Tuikkanen,J., Jussila,S., Cong,L. and Mantsala,P.
 A gene cluster involved in nogalamycin biosynthesis from
 Streptomyces nogalater: sequence analysis and complementation of
 early-block mutations in the anthracycline pathway
 Mol. Gen. Genet. 251 (2), 113-120 (1996)
 96242142
 8668120
 2
 Ylihkonko,K., Hakala,J., Kunnari,T. and Mantsala,P.
 Production of hybrid anthracycline antibiotics by heterologous
 expression of Streptomyces nogalater nogalamycin biosynthesis genes
 Microbiology (Reading, Engl.) 142 (Pt 8), 1965-1972 (1996)
 96349102
 JOURNAL
 MEDLINE
 PUBMED
 8668120
 2
 Ylihkonko,K., Hakala,J., Kunnari,T. and Mantsala,P.
 Production of hybrid anthracycline antibiotics by heterologous
 expression of Streptomyces nogalater nogalamycin biosynthesis genes
 Microbiology (Reading, Engl.) 142 (Pt 8), 1965-1972 (1996)
 96349102
 JOURNAL
 MEDLINE

8760909	PubMed	3	Torkkell, S., Ylihonko, K., Hakala, J., Skurnik, M. and Mantsala, P.
REFERENCE			Characterization of Streptomyces nogalater genes encoding enzymes
AUTHORS			involved in glycosylation steps in nogalamicin biosynthesis
JOURNAL			Mol. Gen. Genet. 256 (2), 203-209 (1997)
MEDLINE			98007868
PUBMED			9349712
AUTHORS			Palmu, K., Hautala, A., Kunnari, T., Hakala, J., Mantsala, P. and Ylihonko, K.
TITLE			Identification and functional analysis of a glycosyltransferase gene (snogZ) and a resistance gene (snorO) from the nogalamicin producer, Streptomyces nogalater
JOURNAL			Unpublished
AUTHORS			Ylihonko, K.P.J.
JOURNAL			Direct Submission
AUTHORS			Submitted (20-FEB-1998) Ylihonko K.P.J., Biochemistry, University of Turku, Arcanum, Vatselantie 2, FIN-20014 Turku, FINLAND
JOURNAL			Revised by [5]
REMARK			6 (bases 1 to 19632)
REFERENCE			Ylihonko, K.P.J.
AUTHORS			Direct Submission
JOURNAL			Submitted (19-OCT-1999) Ylihonko K.P.J., Biochemistry, University of Turku, Arcanum, Vatselantie 2, FIN-20014 Turku, FINLAND
TITLE			On or before Oct 17, 2002 this sequence version replaced
COMMENT			gi:1181156, gi:2916803.
FEATURES			Overlapping sequence: Z48262.
source			Location/Qualifiers
			1..19632
			/organism="Streptomyces nogalater"
			/mol_type="genomic DNA"
			/strain="ATCC 27451"
			/db_xref="taxon:33314"
			complement(152..1618)
			/gene="snogH"
			complement(152..1618)
			/gene="snogH"
			complement(152..1618)
			/function="putative 2,3-dehydratase"
			/codon_start=1
			/transl_table=11
			/product="SnogH"
			/protein_id="CAA12009.1"
			/db_xref="GI:2916804"
			/db_xref="SPTREMBL:O54255"
			/translation="MTKLSAHPAAGHAGVPPDLRLAASARSAGVWLPRTTEVHHWLA TVGHARQVPRVPLAELDGLWRDPATQNIARHSGGFTVEGLDVTIPRAPVPHWEOP IINQEVGLGIVKFEFGVHLCLMQAKLEPCNGNGIQLSPTVQATSNYTRVHGRG VPYLFHQDPARRSILADVRQSGSWFHKKRNRNMVVEEDIDVHDGFCWLSLQ LHELTIVNLNMDTRITVLSCLPFAHAGLAAVAPGADPPFRALVASCADPGYSRHSL GDLSWITDTRVTRTEVTRVPLDGLRSWRRTDERI SHEDGGFDVIGVVRVTRGREV AEWTQPMTEPKAGVVAFLVRPIEGVHLVHARVEPGYDIVEIAPVTOCTPDQSYER LPARAPLFLDVLPARADRVFDAELSEEGGRFVHARNRYLVAETDLAAGFDHPDR WTLQVLLHSHVNIQARSVLACLYGLATAPPR"
			complement(1615..2601)
			/gene="snogG"
			complement(1615..2601)
			/gene="snogG"
			/function="putative dNDP-4-keto-6-deoxyhexose reductase"
			/codon_start=1
			/transl_table=11
			/product="SnogG"
			/protein_id="CAA12010.1"
			/db_xref="GI:2916805"
			/db_xref="SPTREMBL:O54256"
			/translation="MTAGIPRTVAVGATGCVGRHVCDAFTAGHDVVALARRAPAG GHPRFLACDPAGLADLQDARVDVTVNAAGLNLNEDNRYHHVRLVERLL ACTALVAPRLVQIGSVHEXGPVAGDAVDSEPRPVTVYARTKQARFOVLRVRL ADAGCVLIRAVNCGPMTVPSFLGTVTBRLSEAPGEPVELTLSDARDFVGRDL ADAVLRATPRVTGIFNIGRGEALFIRSLVALLDLVAGLPGTGVTKERLQVSRGGD WTRVTRRRRLGWRPRIDRLTSLHDMWRREALTLGLPPATGSHGDGTRFESAL"
			complement(2598..3203)
			/gene="snogF"
			complement(2598..3203)
			/gene="snogF"
			/function="3,5-epimerase"
			/codon_start=1
			/transl_table=11
			/product="SnogF"
			/protein_id="CAA12011.1"
			/db_xref="GI:2916806"
			/db_xref="GOA:O54257"
			/db_xref="SPTREMBL:O54257"
			/translation="MESRTLLVEGAHAFPRVFPDARGCFVSPQQPQPVAAALGRLRP PVAQTNHRSRRGVRGHAFTLTPGIAKYVYCARGRADYVIDIRVSGSPFTGOWDSV VLDNFRMYPVGVGAHFAALEDDETVMYSYLSEYVADNELALSVPDPDLALELPG DTPLLSRDTPAAPTLESDGRLPRYEECVKLEHDLVAG"
			complement(3251..4222)
			/gene="snogB"
			complement(3251..4222)
			/gene="snogB"
			/function="aromatase, polyketide"
			/codon_start=1
			/evidence="experimental"
			/transl_table=11
			/product="SnogB"
			/protein_id="CAA12012.1"
			/db_xref="GI:2916807"
			/db_xref="GOA:O54490"
			/db_xref="SPTREMBL:Q54490"
			/translation="MTAGDTLTTHRTHTVTSAPARRVFDLVADITGWPHTFPPTV HAEYARGASERIRIMATANGEVKAWSHRSLDEGLRVFPROKSOHPVAGMGGEW IIRVDGERSEVILTHDQVEDDPAHVDWTHRAVDNRSGAELAKLPRRNARTARG TLIPTFADEVITGADGVTDYLNKARWQERLPHVAVSLDEPAPGVLGMDTKGPD GTVHTTESVIRICFPQDRIYVQFTMPPLMAVHTGWIRAPAADGHTTIVTSLHTVVVD DSAVTVLGPDATLADARFLERDALGRNSRTTLGFAKYABERARRA"
			complement(4283..5071)
			/gene="snodA"
			complement(4283..5071)
			/gene="snodA"
			/function="polyketide ketoreductase"
			/codon_start=1
			/evidence="experimental"
			/transl_table=11
			/product="SnodA"
			/protein_id="CAA12013.1"
			/db_xref="GI:2916808"
			/db_xref="GOA:O54491"
			/db_xref="SPTREMBL:Q54491"
			/translation="MPDTRGVAVVTGTSIGGLAVKALATGVCASLARSODAV DSTVELRDEKLEVDGRSCDVRSTDDVREFIGFATRFEGVDLVNNAGRNGGGPTAR ITDELWLDVIELNLSVPRVTEVLTGMLAAGRGINIASTGCKQGVILGAPYSA SKHGVTGPTKALGLAKSGITVNAVCBGVETPMGERVGRGYAEVWDTSEAVLERF QAKIPGRYQAQEEVAGLVLYITTTADSIQAALNVCGGLNY"
			5321..6181
			/gene="snoc"
			5321..6181
			/gene="snoc"
			/function="methylase, aklanonic acid"
			/codon_start=1
			/evidence="experimental"
			/transl_table=11
			/product="Snoc"
			/protein_id="CAA12014.1"
			/db_xref="GI:2916809"
			/db_xref="GOA:O54258"
			/db_xref="SPTREMBL:O54258"
			/translation="MDTAEYKQVTAAPDRAAAHYDRLGVETFTMGLRLVDHAPRP GQVLVDGCGRACLPFAARRVAPDGYALGIDIAPMVVEEAAEAERLGMHPHVEVRM DGERPALPERSERVLSYSVIFFEALPALVRYAGLLADGGRIAFSTSPVETDFFP LPPVFTFLPSLLDTPDPPRFEALRRFNLSWLRPDLLEATMRAGGFGVGIIDET VPMVAVSGAWVDWMSHTNGMLLQHLPRDEAQRRLERLITELDAMRRADAPLSIPVP VREVTAVDOR"
			complement(6221..6577)
			/gene="snobB"

```

CDS      complement(6221..6577)
        /gene="snoaB"
        /function="oxigenase, polyketide"
        /codon_start=1
        /transl_table=11
        /product="SnoaB"
        /protein_id="CAA12015.1"
        /db_xref="GI:2916810"
        /translation="MPTRVNDGVDADEVTFVNRFTVHGGAPEFESVFARTAAAFAROP
        GFVHTLLRERDKNSVYNIAMVTDHDAFRALAQPGFLPHATLALSTSEHGLFTA
        RQTLPEGDTTGSGRH"
        complement(6684..8681)
        /gene="snoaA"
        complement(6684..8681)
        /gene="snoaA"
        /function="activator, polyketide"
        /codon_start=1
        /evidence=experimental
        /transl_table=11
        /product="SnoaA"

gene      Query Match      85.0%; Score 17; DB 1; Length 19632;
          Best Local Similarity 100.0%; Pred.No.2.4e+03;
          Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 CGCGGCACCACTCGAC 18
        |||||
Db      19917 CGCGGCACCACTCGAC 18901

RESULT 15
AP003754/c
LOCUS   AP003754              67893 bp      DNA      linear      PLN 02-SEP-2003
DEFINITION
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,
BAC clone:OU1341.A08.
ACCESSION
AP003754
VERSION
AP003754.4 GI:34393697
KEYWORDS
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
SOURCE
1
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oryza sativa japonica (GR3) genomic DNA, chromosome 7, BAC
clone-OU1341.A08
PUBLISHED Only in Database (2001)
2 (bases 1 to 67893)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission
Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kamondai
21-1-2, Tsukuba, Ibaraki 305-8602, Japan
(e-mail:tsasak@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Sep 1, 2003 this sequence version replaced gi:24417154.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), GeneMark.hmm
(http://opal.biology.gatech.edu/GeneMark/), GlimmerM
(http://www.tigr.org/tdb/glimmerm/glmr.form.html), RiceHMM
(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
(http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4
(http://globin.cse.psu.edu/html/docs/sim4.html) gap2
(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
genomic sequence was searched against NCBI Nonredundant Protein
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
sequence database at RGP or DDBJ. Protein homologies of the coding
regions were searched against NCBI Nonredundant Protein database
with BLASTP. ESTs represent the identified cDNA sequences using
BLASTN with the corresponding DDBJ accession no. and RGP clone ID.

```

```

Full-length cDNAs represent the identified cDNA sequences using
BLASTN with the corresponding DDBJ accession no.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with full-length cDNA or
EST homology (covering almost the entire length of partial
sequence) is classified as an 'unknown' protein. A gene predicted
by two or more gene prediction programs is classified as a
'hypothetical' protein according to TRGSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from -21M13 to M13rev of the BAC
clone. This is a partial sequence of OJ1341.A08 clone. This
sequence of OJ1341.A08 clone has an overlap with P0418E08 (DDBJ:
AP004382) clone at 5' end and with OJ1058.C08 (DDBJ: AP003738)
clone at 3' end. The sequence was generated by combining Monsanto
and RGP-Japan sequencing data. Detailed information on overlap and
assembly quality together with annotation of this entry is
available at http://rgp.dna.affrc.go.jp/Genomeseq.html.
FEATURES
            Location/Qualifiers
1..67893
    /organism="Oryza sativa (japonica cultivar-group)"
    /mol_type="Genomic DNA"
    /cultivar="Nipponbare"
    /db_xref="taxon:39947"
    /chromosomes="7"
    /clone="OJ1341.A08"
    complement(702..1800)
    /gene="OJ1341.A08.101"
    complement(join<702..734,830..863,990..1363,1693..>1800))
    /gene="OJ1341.A08.101"
    /note="start and end point are not identified"
    complement(join(702..734,830..863,990..1363,1693..1800))
    /gene="OJ1341.A08.101"
    /note="predicted by GeneMark.hmm etc."
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="BAC83011.1"
    /db_xref="GI:34393698"
    /translation="MKHNLEKKKKGMPFGAISLPGATAAEAVLLCLPAR
    PLCRHAYVCRPWRLSDPFAAAARHAPHLAVRGRLNSYGRGLVDVYVDDA
    SSQDIVKRCACGRCRPAEVSTHGCVALLVDNQLRLVDPVSGAVPVVVDYKISHT
    NVNGRLALWREAMEERLDGV"
    2010..2384
    /gene="OJ1341.A08.102"
    complement(join<2010..>2384)
    /gene="OJ1341.A08.102"
    /note="start and end point are not identified"
    2010..2384
    /gene="OJ1341.A08.102"
    /note="predicted by FGENESH etc."
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="BAC83012.1"
    /db_xref="GI:34393699"
    /translation="MGSPPARAAPPAGPELLVLCRRRALDLRAPATPPSAASPPH
    PHIGSLPAVALPOPSCVAVTVGSPFLPGHPRQAPAPAPHGAEVHGLPSTGFPF
    QTPVSRHGLVTPRFSGVKIH"
    complement(2618..2974)
    /gene="OJ1341.A08.103"
    complement(<2618..>2974)
    /gene="OJ1341.A08.103"
    /note="start and end point are not identified"
    complement(2618..2974)
    /gene="OJ1341.A08.103"
    /note="predicted by GlimmerM etc."
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="BAC83013.1"
    /db_xref="GI:34393700"
    /translation="MGELAAFCRERCIRAEITWAARSHGRPARAGVEEVGNSGVVG

```

```

WNRFPWLPACRFIGSHLSERKSGKAGGMWTELGKANGSGSPRRRGEGGWAGGERR
REGIFARARKGGRGF"
/ gene
complement(3136..4110)
/ gene="OJ1341_A08.104"
complement(3136..54110)
/ mRNA
/ note="start and end point are not identified"
/ gene="OJ1341_A08.104"
complement(3136..4110)
/ CDS
/ gene="OJ1341_A08.104"
/ note="predicted by GeneMark.hmm etc."
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC22532.1"
/ db_xref="GI:24417171"
/ translation="MADGLAEVAPREHPRGAAGQGHVDPRLGRSIERPAGAAROR
GRATGEGGSRGSGRQDGRQWAAALDVAHTGGSGEGRGRGRGHEV
GKAADDERWRFSGWRGSRPRRCRGGOOLGSGKEGLEGEQRSRHGEHQ
PSTTFVRLTGETITVAPGGEVVAEGALVLAWEATAYVEEVGNGNORSIAANV
ATSRWRKWCSGGLGKSGRPGSCDGEVNGASGSRWLKREWRATEAANGGGSG
TRCGSDRDTGRGSRGCVARRGADGGGTARRPFRREAAGARR"
5036..10775
/ gene="OJ1341_A08.105"
join(<5036..5137,5249..5378,5898..5969,6412..7438,
9389..9523,10339..10465,10638..>10775)
/ gene="OJ1341_A08.105"
/ note="start and end point are not identified"
join(5036..5137,5249..5378,5898..5969,6412..7438,
9389..9523,10339..10465,10638..>10775)
/ gene="OJ1341_A08.105"
/ note="contains EST(s): AU0781130(E20319)"
/ codon_start=1
/ product="unknown protein"
/ protein_id="BAC83014.1"
/ db_xref="GI:34393701"
/ translation="MERDPCDGLCRFOIHGTGDCSLPAGGQVKPKVQLCTSGQRN
LYSDRARGVSTRTVLSLTKSTOCTVHMPIAGVLIRPERGLVFTAAGT
LGYDMLPVPIEYVYAIELRVAAALPRPCRLSTPLVVLVLRGALLSWATSP
ADFSPLRHGVAVLSSTAASQASLSSSPFAHQPRPHWLSAIQCLPLLRASPP
VLCVATQRLSLVAPRSPRLVSSDIDGVMVPAVPRVYQHERHTSSRVPG
NCLVQRLSPRSPRSDIASISMTASIPSSARPRAPHTAPRHSVARELLH
SKPCVTRSPRSPRSDIASISMTASIPSSARPRAPHTAPRHSVARELLH
QLPFGYIDHGYSLDGHSLTPYALATSTTAQALRIEHSCKFLQSNWSTSS
SHWLLRNKRCRPHLTAASSSLCNSQSORQRCXDLGLANSSSDHDTED
IETGNEVNYGTTTDLDAETSYDNHISFYVAVHNVANPHITAFDGLPFPKNLHVAV
FISEKFPQSM"
complement(11285..13284)
/ gene="OJ1341_A08.106"
complement(11285..13284)
/ note="probably inactive due to including stop codon(s) in
CDS
pseudogene, gag/pol polyprotein"
/ pseudo
complement(14421..14678)
/ gene="OJ1341_A08.107"
complement(14421..514678)
/ gene="OJ1341_A08.107"
/ note="start and end point are not identified"
complement(14421..14678)
/ gene="OJ1341_A08.107"
/ notes="predicted by GlimmerM etc."
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC83015.1"
/ db_xref="GI:34393702"
/ translation="MGLFRRVIOATLFHAYLQOQVKPRHPKPMRLKWQGGWRHLH
DVTNCTGVRTDEANTGTATPAPSPSCMRREAKTDDDD"
14960..15379
/ gene="OJ1341_A08.108"
<14960..>15379
/ notes="start and end point are not identified"
14960..15379

```

```

/ gene="OJ1341_A08.108"
/ note="similar to Oryza sativa chromosome 4,
OSJNBa0050F15.5"
/ codon_start=1
/ product="hypothetical protein"
/ protein_id="BAC83016.1"
/ db_xref="GI:34393703"
/ translation="MTKNLLBRADDDEHRLVALRRLTOARRRDVPGNSDEMAVD
VAGSADLARAPSSGTPAGVLEEQPADEEEDGNDVDESETSVMSRGGCDREGDNVQR
LGDITRRRGIEVAGAEARTSQQVDVSPATDVSDERE"
complement(18618..19148)
/ gene="OJ1341_A08.109"
complement(<18618..>19148)
/ gene="OJ1341_A08.109"

Query Match 85.0%; Score 17; DB 8; Length 67893;
Best Local Similarity 100.0%; Pred. No. 2.2e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGCGCCACCACTCGAC 18
Db 33523 CGCGCCACCACTCGAC 33507

RESULT 16
AC135598 137312 bp DNA linear HTG 16-OCT-2003
Oryza sativa (japonica cultivar-group) chromosome 3 clone
OSJNB0021K20, ** SEQUENCING IN PROGRESS **, 2 ordered pieces.
AC135598
AC135598.3 GI:37693584
HTG; HTGS PHASE2; HTGS ACTIVEFIN.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 137312)
Buell, R., Hsiao, J., Zismann, V., Moffat, K.M., Hill, J.,
Gansberger, K., Burgess, S., Jarrahi, B., Shvartsbeyn, M., Brenner, M.,
Ciecko, A., Pai, G., Vanaken, S., Hansen, C., Utterbach, T.,
Feldblum, T., Khalak, H.G., Yuan, Q., Quackenbush, J., White, O.,
Salzberg, S. and Fraser, C.
Oryza sativa ssp. japonica cv. Nipponbare OSJNB0021K20 BAC genomic
sequence
Unpublished
2 (bases 1 to 137312)
Buell, R.
Direct Submission
Submitted (19-OCT-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Buell, R.
Direct Submission
Submitted (16-OCT-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Oct 15, 2003 this sequence version replaced gi:24418097.
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 68692: contig of 68692 bp in length
* 68693 68792: gap of unknown length
* 68793 137312: contig of 68520 bp in length.
Location/Qualifiers
1. 137312
/ organism="Oryza sativa (japonica cultivar-group)"
/ mol_type="genomic DNA"

```

FEATURES
source


```

/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="3"
/clone="OSJNB0021K20"

Query Match      85.0%; Score 17; DB 2; Length 137312;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
2 CGCGCACCACCTCGAC 18
|||||
Db 110026 CGCGCACCACCTCGAC 110042

RESULT 17
CNS08C9S      158261 bp      DNA      linear      PLN 21-NOV-2003
LOCUS
DEFINITION
Oryza sativa chromosome 12, . BAC OSJNBa0040E18 of library OSJNBa
from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza
sativa (rice), complete sequence.
ACCESSION
AL772414
VERSION
AL772414.3 GI:24430251
KEYWORDS
HTG.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Ehrhartoideae; Magnoliophyta; Liliopsida; Poales; Poaceae;
Oryza.
REFERENCE
1 (bases 1 to 158261)
Choisne N., Orjeda G., Cattolico L., Demange N., Wincker P.,
Segrens B., Pelletier E., Scarpelli C., Salanoubat M.,
Weissenbach J. and Quetier F.
Oryza sativa chromosome 12 sequencing
Unpublished
TITLE
Oryza sativa chromosome 12 sequencing
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 158261)
Genoscope.
Direct Submission
Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
On Oct 30, 2002 this sequence version replaced gi:23304718.
Center: Genoscope / Centre National de Sequencage
Contact: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : OJ112 B07 (AC=AL731753)
Downstream BAC (overlapping the SP6 end) : OSJNBa0057C04
(AC=AL954159) ----- Finishing boundaries
FINISHED SEGMENT STARTS AT BASE 46454
FINISHED SEGMENT ENDS AT BASE 146009
-----
FEATURES
source
Location/Qualifiers
1. 158261
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="12"
/clone="OSJNBa0040E18"
/clone_lib="OSJNBa"

Query Match      85.0%; Score 17; DB 8; Length 158261;
Best Local Similarity 100.0%; Pred. No. 2.1e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
2 CGCGCACCACCTCGAC 18
|||||
Db 110026 CGCGCACCACCTCGAC 110042

RESULT 18
AE016783/c
LOCUS
DEFINITION
Pseudomonas putida KT2440 section 10 of 21 of the complete genome.
ACCESSION
AE016783 AE015451
VERSION
AE016783.1 GI:26557027
KEYWORDS
Pseudomonas putida KT2440
Pseudomonas putida KT2440
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
SOURCE
1 (bases 1 to 308015)
Nelson K., Paulsen I., Weinal C., Dodson R., Hilbert H., Fouts D.,
Gill S., Pop M., Martins Dos Santos V., Holmes M., Brinkac L.,
Beanan M., DeBoy R., Daugherty S., Kolonay J., Madupu R.,
Nelson W., White O., Peterson J., Khouri H., Hance I., Lee P.,
Holtzapple E., Scanlan D., Tran K., Moazzaz A., Uterback T.,
Rizzo M., Lee K., Kosack D., Moestl D., Wedler H., Lauber J.,
Hohseil J., Straetz M., Heim S., Kiewitz C., Eissen J., Timmis K.,
Duesterhoft A., Tummler B. and Fraser C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 793-808 (2002)
2 (bases 1 to 308015)
Nelson K., Paulsen I., Weinal C., Dodson R., Hilbert H., Fouts D.,
Gill S., Pop M., Martins Dos Santos V., Holmes M., Brinkac L.,
Beanan M., DeBoy R., Daugherty S., Kolonay J., Madupu R.,
Nelson W., White O., Peterson J., Khouri H., Hance I., Lee P.,
Holtzapple E., Scanlan D., Tran K., Moazzaz A., Uterback T.,
Rizzo M., Lee K., Kosack D., Moestl D., Wedler H., Lauber J.,
Hohseil J., Straetz M., Heim S., Kiewitz C., Eissen J., Timmis K.,
Duesterhoft A., Tummler B. and Fraser C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
Location/Qualifiers
1. 308015
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
122..2422
/gene="cti"
/locus_tag="PP2376"
122..2422
/gene="cti"
/locus_tag="PP2376"
122..2422
/gene="cti"
/locus_tag="PP2376"
/locus_tag="PP2377"
/locus_tag="PP2377"
/product="esterified fatty acid cis/trans isomerase"
/protein_id="AA067989.1"
/db_xref="GI:24983931"
/translation="MVHRIAGAFALLISGAVFGQAPQSSPAISYTRDIQPIFTBKV
ACHACNDAACQLKLESPGAVRGASKVPYQGERSKAVPTTFLFYDAKSEOWRKKGF
VSLDNGQGRALMARMLGHKHTPLTNKLPETVLGLSNMCPLEHEDAYAGA
HPKGMPLAVTGLTDKEDTMRRLAAGAPVEYQIQPSAERAOIAEWEELLNPGS
TEALVGRMDYEHFLAHFHAGGEGHFFQVWRSTPSGKFPVDIIATRRPNDPPGTF
YVRLIPVQGVIVHKHTIYPMGPQKLVKQKLFYAGDHAALPGYGRHRANPETF
KYGREATPLLMGQIDVGVSVLSLHAYRKNDYKLRREAYAEAPRWSTLWAG
EAPAVARYQFMDNAEYFVTFIRGPVCRGQIATDVRDNFWALQEFADFRTIDA
NDNALLSIFRSDASVTGKLVGDPVLTWLPDYPLFERTYVQLAVNFDVYGNVSHOL
OTELYEDLIRNGAEVNFRLMPADOKKALGDMYONSGKVMMDYEDIDTDPGSIK
LDPRNPKRDFGLKLLQRTGSLNAADPPINRCGAGFCSPQMSSEFRNAEQSLRSLVR
PAAGLKVINQLPEATMLRIEGDQGRQVYSLLRNASHNAFLLGCAEYRQGLDTLT
LYPGVLSYVPNFIFNIPYQDPEFVEDMEYAKDDAAKFERIVMRGVRSHDPAFWRYF
HDLNSYIKETPEVEAGVLDNRYENL"
2500..3651
/locus_tag="PP2377"

```

```
CDS
2500...3651
/locus_tag="PP2377"
/translation="similar to GP:5257179, and GP:5257179; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acyltransferase Act, putative"
/protein_id="AAN67990.1"
/db_xref="GI:24983932"
/translation="MQPAGPAPTGTGALLRLALASALDTGLNAPPSAGYMLYSL
QALLAFAMVVCHHEMGIFFDHATGPIGOLLDRGAVGVDIFPVISGLVILYSTRD
KTVPRPOLLARLIVPAVWFYALMALILLAPSQMHPQSDWRHLFLSLLFIAPAE
NPGVGLPTLVNGWTINPFYLLPGLAFVLRQHLLVYVLAALLVSEVLGRGV
VSRFNNDDIIYEFLLIGVGLYRRGLRQWPLALLGVAGNATVCLDSORLJHW
GLPSAMIVAFVMEPEFQGNRLKALGDCSYVLIHVLVLYAGWAFSQRLLHNPFL
TFALCPVSGILMSWFSYQWLERGLYRRMQAWLAQGPAPAYALSRVNC"
3710...4294
/gene="yhgi"
/locus_tag="PP2378"
3710...4294
/gene="yhgi"
/locus_tag="PP2379"
/translation="similar to GB:M68840, GB:X60819, SP:P21397,
PID:187351, PID:187353, and PID:187355; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="yhgi protein"
/protein_id="AAN67991.1"
/db_xref="GI:24983933"
/translation="MSAITIDAAHDYLDLSSKQNTPGIGIRIFITPGTQVAYETCI
AYCPGEKPDPTAVGLKSFAYLDVAVSPELEADYVYATDRMGOLITKAPNAKVP
WYNDSPTNERINYLQTEINPGLASHGQVSLVDVDDGIANVLQFGGCGCGQADV
TLKGIERTLLERPELQVGRVDVTHDSQENAY"
4428...5099
/locus_tag="PP2379"
/locus_tag="PP2379"
/locus_tag="PP2379"
/translation="similar to GP:3298363, SP:P04127, GB:X02921,
SP:P02972, SP:P04740, GB:M68060, and GB:X03391; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="Scol/SenC family protein"
/protein_id="AAN67992.1"
/db_xref="GI:24983935"
/translation="NPRFDRPVRACFPVKDGFMDLLTRAVVAGMVLGLLLA
GCSFARGDEFKIGKMSNEILGRKFSKADPGQNVRTLSFFYSGSPMVFPGFTQCPAVC
PTTLARAQIRKLGRDRDLQVVFITLDPEDTPEVLDAYKAFDPTFTALGTPE
EIAAVAKFKVFEKVPAGDTYTIHSSTSYVYDTRGLRLSLGSLNAKEADLVY
LMEIC"
5100...5582
/locus_tag="PP2380"
5100...5582
/locus_tag="PP2380"
/translation="similar to GB:D13178, SP:Q08751, PID:398125,
GB:D13178, SP:Q08751, and PID:398125; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN67993.1"
/db_xref="GI:24983936"
/translation="MSMQPIKREGAALVLMGLALPALAQTTVSDAWVRASVPHQOSTG
APMVLTAQSDSKLVGVAPKTVQVHMTMNGVMGMRVKATIELPAGKAVITLDPNG
LHVLMLGLHNQVKEGDKVPLTLTIEDAKGETLEVQAVRALNADAGGGHMHNNH"
complement(5632...5874)
/locus_tag="PP2381"
complement(5632...5874)
/locus_tag="PP2381"
/codon_start=1
/transl_table=11

gene
/translation="conserved domain protein"
/protein_id="AAN67994.1"
/db_xref="GI:24983937"
/translation="MEDPAMGERELTLLSLMNOFQAQLSSACKIEADWIDROGDVPA
AGKIRASLKALEADEAQVRKTLTSLDRPLPRFRS"
6229...7533
/locus_tag="PP2382"
6229...7533
/locus_tag="PP2382"
/translation="similar to GB:M11189, SP:P04190, SP:P14488, and
PID:142604; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transporter, putative"
/protein_id="AAN67995.1"
/db_xref="GI:24983938"
/translation="MAISIGSKECPYESAIOVRSADVARKYPLMLTLEINFGTSLSL
GYSKGGEFAFEAFVAIVISITADMLLIGIAVFFRILPSVVFSPGLGWLADNANK
KHTLAEIVTKGLLSPALFSPVFTVLLVGLAMVMTALECLYVTPFRFPDLV
KDELATVNSGIQVIEDAASILGLVFSACVLLSREATFLFCMLIVSAICITLT
PARCSVQAFDCCALIRDAAGSVSQLRAGNAPLFAVIGCTTLCAMFATSVIRFILPAS
VLEHFASEAAGVGFVPSLLAAGVVGCMLYTRFENRTAGLVIRYMLYVGLGFFFAAVA
LOFNWMLFLLILVGFICAFVDAIVITNIOCLREHEVGRNFSLYITAVMSOAVSG
LIASLVFVLGAPATPIWTLMLLIAPVRWNMKRGPVDVDSRT"
complement(7582...8427)
/locus_tag="PP2383"
complement(7582...8427)
/locus_tag="PP2383"
/translation="identified by match to PFAM protein family HMM
PF00165"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, AraC family"
/protein_id="AAN67996.1"
/db_xref="GI:24983939"
/translation="MSPPLREQTHLWQAPALGDVEMLHARYQORFAPHVHEGVYFTV
IESGAORFWRHSGSEHLAPVGMVLINDELHTGATAHEAGRWYRGFYDEHRSVGVLD
EQLRHGMPSFKSDIHDPALAFSQLHQLSEASALQOQTAWROALVQRHG
CUALPSAQHPEPLAVARRELLSQLADPPSLEAALAAVNLSPHFARVFRQATGLPP
HAWLKORLARAREMLKEGLAASQVAFDLGADQSHLSRQKQAYGVTPGAYRQACVH
SALRA"
complement(8424...8738)
/locus_tag="PP2384"
complement(8424...8738)
/locus_tag="PP2384"
/translation="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN67997.1"
/db_xref="GI:24983940"
/translation="MTWILIFAMGAIVFLNRYAFLEPRLPURLSSNAQFLGFAVPGM
LTAICGPIIFMPDHQNLNLLNPNYLLGLSLVAIALVLTSLLSMLVSLIFFLLRSW
LA"
complement(8735...9427)
/gene="azlC"
/locus_tag="PP2385"
complement(8735...9427)
/gene="azlC"

Query Match 85.0%; Score 17; DB 1; Length 308015;
Best Local Similarity 100.0%; Pred. No. 1.9e-03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GCCGCACCCCTCGACT 19
|||||
Db 46664 GCCGCACCCCTCGACT 46648

RESULT 19
AF087589/c 276 bp DNA linear PLN 08-FEB-2000
LOCUS Saxifraga marginata internal transcribed spacer 1, partial
DEFINITION
```

```

sequence.
ACCESSION AF087589
VERSION AF087589.1 GI:5814034
SOURCE Saxifraga marginata
ORGANISM Saxifraga marginata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Saxifragales; Saxifragaceae; Saxifraga.
REFERENCE 1 (bases 1 to 276)
AUTHORS Conti, E., Soltis, D.E., Hardig, T.M. and Schneider, J.
TITLE Phylogenetic relationships of the silver saxifrages (Saxifraga,
sect. Ligulatae haworth): implications for the evolution of
substrate specificity, life histories, and biogeography
Mol. Phylogenet. Evol. 13 (3), 536-555 (1999)
JOURNAL 20088704
MEDLINE 10620412
PUBMED
REFERENCE 2 (bases 1 to 276)
AUTHORS Conti, E., Soltis, D.E., Michael, H. and Jason, S.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-1998) Museum and Department of Biology and
Wildlife, University of Alaska-Fairbanks, 907 Yukon Drive,
Fairbanks, AK 99775-6960, USA
FEATURES
source 1..276
Location/Qualifiers
misc_RNA /organism="Saxifraga marginata"
/db_xref="taxon:102731"
<1..>276
/product="internal transcribed spacer 1"
ORIGIN
Query Match 84.0%; Score 16.8; DB 8; Length 276;
Best Local Similarity 90.0%; Pred. No. 4.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCGCACCACTCGACTT 20
Db 113 GCGCGCGCACCGCCCGACTT 94
RESULT 20
CXSTPLUC1/c 420 bp DNA linear BCT 22-AUG-1996
LOCUS
DEFINITION C.xyli DNA for strong promoter (420 bp).
ACCESSION X87993
VERSION X87993.1 GI:1321800
KEYWORDS hairpin loop; luciferase; repeated sequence; strong promoter.
SOURCE Leifsonia xyli
ORGANISM Leifsonia xyli
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococccineae; Microbacteriaceae; Leifsonia.
REFERENCE 1
AUTHORS Haapalainen, M., Karp, M. and Metzler, M.C.
TITLE Isolation of strong promoters from Clavibacter xyli subsp.
cynodontis using a promoter probe plasmid
Biochim. Biophys. Acta 1305 (3), 130-134 (1996)
JOURNAL 96180315
MEDLINE 8597597
PUBMED
REFERENCE 2 (bases 1 to 420)
AUTHORS Haapalainen, M.L.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-1995) M.L. Haapalainen, University of Turku,
Department of Biology, Biocity A 6, FIN-20520 Turku, FINLAND
FEATURES
source 1..420
Location/Qualifiers
misc_structure 271..310
/db_xref="taxon:1575"
/sub_species="cynodontis"
/strain="plasmidless, a natural isolate from Florida,
U.S."
/mol_type="genomic DNA"
/organism="Leifsonia xyli"

```

```

/note="hairpin
putative"
repeat_unit 275..282
/rpt_type=DIRECT
repeat_unit 317..324
/rpt_type=DIRECT
-35_signal 369..374
-10_signal 392..397
mRNA 401..>420
411..>420
ORIGIN
Query Match 84.0%; Score 16.8; DB 1; Length 420;
Best Local Similarity 90.0%; Pred. No. 4.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCGCACCACTCGACTT 20
Db 44 GCGCGCGCACCACTCGACTT 25
RESULT 21
AX122356/c 858 bp DNA linear PAT 11-MAY-2001
LOCUS
DEFINITION Sequence 2272 from Patent EP1108790.
ACCESSION AX122356
VERSION AX122356.1 GI:14039309
KEYWORDS Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 2272 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source 1..858
Location/Qualifiers
misc_RNA /organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
ORIGIN
Query Match 84.0%; Score 16.8; DB 6; Length 858;
Best Local Similarity 90.0%; Pred. No. 3.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCGCACCACTCGACTT 20
Db 36 GCGCGCGCACCACTCGACTT 17
RESULT 22
BD164473/c 858 bp DNA linear PAT 17-JAN-2003
LOCUS
DEFINITION Novel polynucleotide.
ACCESSION BD164473
VERSION BD164473.1 GI:27870285
KEYWORDS JP 2002191370-A/2272.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 858)
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotide
JOURNAL Patent: JP 2002191370-A 2272 09-JUL-2002;
KYOWA HAKKO KOGYO CO LTD
COMMENT OS Corynebacterium glutamicum
PN JP 2002191370-A/2272
PD 09-JUL-2002

```

PF 15-DEC-2000 JP 2000405096
 PI SATOSHI NAKAGAWA,HIROSHI MIZOGUCHI,SEIKO ANDO,MIKIO HAYASHI,
 PI KEIKO OCHIAI,
 PI HARUHIKO YOKOI,NAOKO TATEISHI,AKIHIRO SENO,MASATO IKEDA,AKIO
 PI OZAKI
 PC C12N15/09,C12N15/09,C07K14/34,C07K16/12,C07K16/40,C12M1/00,PC
 C12N1/15,
 PC C12N1/19,C12N1/21,C12N5/10,C12N9/00,C12N9/02,C12P7/40,C12P13/13/
 PC 04,C12P13/08,
 PC C12P19/00,C12P19/34,C12P21/02,C12Q1/37,C12Q1/68,G01N33/53,PC
 G01N33/566,
 PC G01N33/569,G01N33/68,G01N37/00//C12P21/08,(C12N1/21,C12R1:15),
 PC (C12N1/21,C12R1:13),(C12N1/21,C12R1:01),(C12P13/08,C12R1:15),
 PC C12N15/00,
 PC C12N15/00,C12N15/00
 CC Novel polynucleotide
 FH key Location/Qualifiers
 FT source 1..858
 FT Location/Qualifiers
 1..858
 /organism="Corynebacterium glutamicum".
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"

ORIGIN

Query Match 84.0%; Score 16.8; DB 6; Length 858;
 Best Local Similarity 90.0%; Pred. No. 3.8e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 DB 36 GCGCGCACCACTCGACTT 17

RESULT 23
 AK108178
 LOCUS
 DEFINITION
 AK108178 968 bp mRNA linear PLN 24-JUL-2003
 Oryza sativa (japonica cultivar-group) cDNA clone:002-140-A08, full
 insert sequence.
 ACCESSION
 AK108178 GI:32993387
 VERSION
 FLI CDNA; oligo capping.
 KEYWORDS
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
 1 The Rice Full-Length cDNA Consortium, National Institute of
 Agrobiological Sciences Rice Full-Length cDNA Project Team,
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
 Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
 Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
 Ohtsuki, K., Shisshiki, T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group, Ootomo, Y., Murakami, K.,
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
 Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
 Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
 Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
 Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
 Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
 Hara, A., Hashizume, W., Hayashida, K., Imotani, K., Ishii, Y., Itoh, M.,
 Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Oota, Y., Ooka, H.,
 Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Yoshino, M. and Hayashizaki, Y.
 Collection, mapping, and annotation of over 28,000 cDNA clones from
 japonica rice
 Science 301 (5631), 376-379 (2003)
 JOURNAL
 MEDLINE
 PUBLISHED
 12869764
 REFERENCE
 2 (bases 1 to 968)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,

Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
 Hori, F., Hotta, I., Iida, J., Ikeda, J., Ikeda, R., Imamura, K.,
 Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
 Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
 Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
 Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
 Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
 Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
 Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,
 Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
 Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,
 Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
 Shinagawa, A., Shiraki, T., Shisshiki, T., Sogabe, Y., Sugano, S.,
 Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
 Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
 Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
 Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
 Yoshimura, A.
 Direct Submission
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kanondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)
 This clone is one of the 28K full-length cDNA clones from japonica
 rice.
 URL: http://cdna01.dna.affrc.go.jp/cDNA/
 NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K.,
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shisshiki, T. and
 Yamamoto, M.
 PAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y.,
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
 Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
 Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,
 Yoshimura, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
 Yoshimura, A., Matsubara, K. and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
 Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
 Yasunishi, A. and Hayashizaki, Y.
 Location/Qualifiers
 1..968
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="002-140-A08"

FEATURES
 source

ORIGIN

Query Match 84.0%; Score 16.8; DB 8; Length 968;
 Best Local Similarity 90.0%; Pred. No. 3.8e+03;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 DB 672 GCGCGCACCACTCGACTT 691

RESULT 24

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1755)
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Dresnek, D., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Kronmiller, B., Li, P., Liao, G., Miranda, A., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M., and Celniker, S.

TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA

COMMENT Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES
Source
1..1755
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="48F1-48F1"
1..1755
/gene="guf"
/note="alignment with genomic scaffold AE003823"
/db_xref="FLYBASE:FBgn0014184"
1358..1564
/gene="guf"
/note="Longest ORF"
/codon_start=1
/product="RE14453p"
/protein_id="AAL48693.1"
/db_xref="GI:17945272"
/db_xref="FLYBASE:FBgn0014184"
/translation="MFSLINIVFMFFLYILCLSLYKVCNPLFALKNIQVIFH FLYABSDENGEVLEKDKGNAYA"

ORIGIN
Query Match 84.0%; Score 16.8; DB 3; Length 1755;
Best Local Similarity 90.0%; Pred. No. 3.6e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCCACCTCGACTT 20
DB 244 GTCCGCACCTCTCGACTT 263

RESULT 27
DMU29529
LOCUS Drosophila melanogaster gutfeeling (guf) mRNA, complete cds.
DEFINITION U29529
ACCESSION U29529.1 GI:1143044
KEYWORDS

SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 2163)
AUTHORS Salzberg, A., Golden, K., Bodmer, R. and Bellen, H.J.
TITLE gutfeeling, a Drosophila gene encoding an antizyme-like protein, is required for late differentiation of neurons and muscles
JOURNAL Genetics 144 (1), 183-196 (1996)
MEDLINE 97032936
PUBMED 8878684

REFERENCE 2 (bases 1 to 2163)
AUTHORS Salzberg, A., Golden, K., Bodmer, R. and Bellen, H.J.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1995) Adi Salzberg, HHMI, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA

FEATURES
Source
1..2163
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/chromosome="2"
/map="48E"
1..2163
/gene="guf"
1261..1821
/gene="guf"
/standard_name="gutfeeling"
/note="Method: conceptual translation supplied by author; similar to human ornithine decarboxylase antizyme, encoded by Genbank Accession Number U09202"
/codon_start=1
/protein_id="AAB49330.1"
/db_xref="GI:1143045"
/translation="MLKCVNSFRVGYSRKSSVDSAGGSGLYBASSRSLSSQSDCS LESOPDIHSLCSDDDCOEVLROILQDQVOITIKLHVTEQYTNWNTILNPVNNLLY VALPKDLPPAGSKOTFISLEPAEFKLEVDGIVMVPKQDQDRLRLIEAFLFMGFEPL SRKAQAPPAINDNENTYFLYSIEE"

ORIGIN
Query Match 84.0%; Score 16.8; DB 3; Length 2163;
Best Local Similarity 90.0%; Pred. No. 3.6e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCCACCTCGACTT 20
DB 1082 GTCCGCACCTCTCGACTT 1101

RESULT 28
AE011952
LOCUS Xanthomonas axonopodis pv. citri str. 306
DEFINITION the complete genome.
ACCESSION AE011952 AE008923
VERSION AE011952.1 GI:21109405
KEYWORDS Xanthomonas axonopodis pv. citri str. 306
SOURCE Xanthomonas axonopodis pv. citri str. 306
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xanthomonas.

REFERENCE 1 (bases 1 to 11366)
AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, I.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J., Chambergo, F., Cipina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Curcio-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Kura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, P.F., Setubal, J.C. and
 Kitajima, J.P.
 Comparison of the genomes of two Xanthomonas pathogens with
 differing host specificities
 Nature 417 (6887), 459-463 (2002)
 22022145
 12024217
 2 (bases 1 to 11366)
 da Silva, A.C.R., Ferro, J.A., Reinach, P.C., Farah, C.S., Furlan, L.R.,
 Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
 Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
 Camargo, L.E.A., Canarotte, G., Cannavan, F., Cardoso, J.,
 Chambergo, F., Chapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
 Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,
 Ferreira, R.C.C., Ferro, M.T., Formighieri, E.F., Franco, M.C.,
 Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
 Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locati, E.C., Machado, M.A.,
 Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
 Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
 Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
 Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
 Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
 Santos, M., Truffi, D., Tsai, S.M., White, P.F., Setubal, J.C. and
 Kitajima, J.P.
 Direct Submission
 Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
 São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900,
 Brazil

FEATURES

Location/Qualifiers

1..11366
 /organism="Xanthomonas axonopodis pv. citri str. 306"
 /mol_type="genomic DNA"
 /strain="306"
 /db_xref="taxon:190486"
 /note="pathovar: citri"
 130..1971
 /gene="bga"
 /note="synonym: XAC3084"
 130..1971
 /gene="bga"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="beta-galactosidase"
 /protein_id="AA037929.1"
 /db_xref="GI:21109406"
 /translation="MLRTLPIALVLAFAIPITGTAETRPWFNFGTQGTQFVRDCK
 PYLLSGAIHFQIPRIAYWKDLQKARALGLNVTYVFNVLVEPQOQDFRSGHNDV
 AAFVREAAQGLNVLPGPYCAEWAGGYPAWLGKGNIRVSRDPRFLAASQAVL
 DALANQVPLNNGGPIIAVOVENYGSYADDAHYADNRMHYVAGDFAKLLFTSD
 GADMLANGTLPDIAVNFAPGSAFDFKLIKFRPDQPRVGEYAWGHWGKPHYA
 AIDTAQAEFEWLRQCHGANLYMFIGTSGFGMGNANFNQNSDHYAPQTTSDYD
 AILDEAGHPFPKALMDAIAIVTGVQFPALPAPIAITLPTFLRESASLWDLNLPAP
 IAITDPQMGQCYILYRTVTTPGRKGLYLDGVRDVARVYVQRPVGSVERRL
 QVSLDVEIPAGQHTLVLVENSGRINYGTRMADGRAGLVDPVLDNRLQITGQAFPL
 PMPTDSIRGWTRKAVGAPAFRGTLIGITPTDYLDMRAFSGKGFANWGNLGRHN
 IGPOTALYFAPFQKGNQNTVVFDLDDVAIVPSVRLKQVWLKPKX"
 complement (2021..4945)
 /gene="XAC3085"
 complement (2021..4945)
 /gene="XAC3085"
 /note="putative; ORF located using Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA037930.1"
 /db_xref="GI:21109407"
 /translation="MGVLLRGGSTAGLLCRRVLARQLRQAIHAAACILPIRRGQTQ

WRAFGINRTRWAGAFVYGIACSSQQDWMERTGAPSLCTGWSLLNRVORSTQHL
 VTRDCAAVTPPAESPDQRGWPLGNALQTRPLRTHRDESSRRVHHGDANARVA
 EAEYDVFVAKYVADPIHGTADTAELINDIPRRGPPEHSAEAGSAHPFAHPAD
 HAAADPHSPFPQTAHPAQHAHAPHAPTPEGHQAGHAGHAPADPAQVCHAKEGV
 LHALAEAGAPAGFDLMSGLISGAMLPLSLAIYAAVYKREVAEQTHLRQERRL
 LEQALOSALGTAPASGAANAHAHALSEAIDTNAVLOARNRNDGAIATASMASAGVI
 FTKASELIGIOGLAIASKSANAAGLVGHSAAAGAAAGAGIAGSFVLAPLASVAATA
 LGSTSHQRERRARVAVDVRGRVFLQOLPESELSPGAQRYQYFVSTKLGKIDRFPAH
 RFNWNKGVVGGITTYTATLTGKVSAAVAGATAAPVGTGLIIGAGLLGAATMGV
 GSHQFLLAGKQKRYRYQTEDMPCVDALLAVADLLPAREAESAAVQETPNVADQLQ
 PTALGVALTPTVAESAGSEEDIVAEADPGLEPKHROGSAAGAQAPMPQDGAAD
 ARAAGNAVPOGFELRSALYACIDGQEKALDAFLQSGADDLQKLPKPRSTDSQKGS
 GRAAHPLSRVRNALRAAANYGRAVTVGKPHQARGKAMRSHAKHSAVLITETLAOW
 LOTPSQKSOIAYMCCLLCTKYLDAKLGAQADLPQDVHDDPHSHSQGTPTVATVDRRA
 PERLLAQHQVTRDQTRAVSLMLQELMAEQELANPDARSQARAARMPGLQRL
 TRVLSNAIVPOEGVAGFARFCMKQARHTTHVRGTLTATEVQAARIREQAAATPAS
 TT"
 5178..5378
 /gene="XAC3086"
 5178..5378
 /gene="XAC3086"
 /note="putative; ORF located using Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA037931.1"
 /db_xref="GI:21109408"
 /translation="MSSSTVAQESVSAVAHTQNAFAEQIHSKPMGSAEAAHMTFLC
 QTESAKKPTLLPQMRDDDA"
 complement (5436..5765)
 /gene="rebB"
 /note="synonym: XAC3087"
 complement (5436..5765)
 /gene="rebB"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="rebB protein"
 /protein_id="AA037932.1"
 /db_xref="GI:21109409"
 /translation="VAFPTAVNPQITDAVTQTNVVGGEAPAVANSSIIQSMHSSGI
 LFNVAQAQQQODTLQAATNGVIQTYGVDTGACADVENVAQGAADNLNLSLTVL
 RSFDFAK"
 5784..6089
 /gene="XAC3088"
 5784..6089
 /gene="XAC3088"
 /note="putative; ORF located using Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA037933.1"
 /db_xref="GI:21109410"
 /translation="MDTRFVLCRPIGAKNASRLCSPHRRRTATPRSLCHHDSGRRR
 RNSTUALFRFSARRNLGVIVRAGAVRLSHRCTEGFTAAMLDHGRVHSRARVPGAIS"
 6226..7386
 /gene="XAC3089"
 6226..7386
 /gene="XAC3089"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA037934.1"
 /db_xref="GI:21109411"
 /translation="MHPRLSLRLVASLAASI.GAVVAQASPPSPIMLSEFIADP
 APTQAHASTLLETRDGHLLAAWFCGEHAGADVGINLAQRPQWQTPRIADGAHA
 GVDGAVPAWPFVIFQSATGPVQLIYKLGPNRQWNGRLSLTDCAGSPFPAQIAG
 ILGPIKNKPVQLPNRILAPSSSDRGWRAHLESDDDGHWQRWQRWQRLKAIQ
 PELLMPGRLOALGRSQNKLFSTFSLDGLHQWPMYLLLEVENSTGSDAVYLHGR
 ALLVYNPAIAGKMDWDRGTGLAVASNDGVHQRVLTLEDSAGDEFSPAVITQTRDGL

gene

complement (2021..4945)
 /gene="XAC3085"

CDS

complement (2021..4945)
 /gene="XAC3085"

gene

complement (2021..4945)
 /gene="XAC3085"

CDS

complement (2021..4945)
 /gene="XAC3085"

```

VHVSYTKRRRIKHVLDPTFRGECSPQAPVPTMLPERCTTACAHADPSS"
7865..9358
/gene="XAC3090"
CDS
7865..9358
/gene="XAC3090"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="leucin rich protein"
/protein_id="AA037935.1"
/db_xref="GI:21109412"
/translation="MPTVIGASPLRRSPAIRPYEDVLVQWRHVSADRNWHSAWROA
NSNNPEVRRTRALKATDLELLEATAPERVALEKSVLPQFPDQAFRUSHQHMTH
DAAGMELPDTMQFAGLETTLARNPLRALPASLSRLRELISRACPELTELPEH
LASTDASGEHQVNLQSLRLERTGITSLPASIANLQNLKSLKINSPLSALGPAIHQ
LPKLELDLRCGTALRSYPIFGSAPLRLIKDCNLLTLPLDIHRLTQLEKLDLR
GCNLSRLPSLIARLPAACIILVPHQAQLDQHVALEKVALERIEDTQAMLS
TVIEERNPPLFEGAPSYLPEKPPGCTTTEGIEIPLQMLEESDPEHLOVSDMAGP
SPRIESEEGLRLHYTNVSNKQAQSAHLGIVDHLGCFVYHERASQDASLAKAVOM
WKTRELIYNAHPQDRTSPDFTLHPHQVSDADDEQOTSPAPSRDQ"
complement(9460..10257)
/gene="cutC"
/note="synonym: XAC3091"
complement(9460..10257)
/gene="cutC"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="copper homeostasis protein"
/protein_id="AA037935.1"
/db_xref="GI:21109413"

Query Match      84.0%; Score 16.8; DB 1; Length 11366;
Best Local Similarity 90.0%; Pred. No. 3.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCTCGACTT 20
DB 9955 GCGCGCACCTCGACTT 9974

RESULT 29
AC015208
LOCUS
DEFINITION
AC015208
AC015208.1 GI:6436127
VERSION
HTG; HTGS PHASE2.
KEYWORDS
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
Adams,M. and Venter,J.C.
Direct Submission
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM.10211252 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..64195
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
source
1..64195
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

```

```

Query Match      84.0%; Score 16.8; DB 2; Length 64195;
Best Local Similarity 90.0%; Pred. No. 2.7e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCTCGACTT 20
DB 41235 GCTCCGCACATCTCGACTT 41254

RESULT 30
AP004394/c
LOCUS
DEFINITION
AP004394
AP004394.1 GI:17402600
VERSION
HTG; HTGS PHASE2.
KEYWORDS
Oryza sativa (japonica cultivar-group)
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzoae; Oryza.
1
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
clone:QJ1118.C06
Published Only in Database (2001)
2 (bases 1 to 98286)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (05-DEC-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7488
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..98286
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="QJ1118.C06"

ORIGIN
source
1..98286
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="QJ1118.C06"

Query Match      84.0%; Score 16.8; DB 2; Length 98286;
Best Local Similarity 90.0%; Pred. No. 2.6e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCTCGACTT 20
DB 33269 GCGCGCACCTCGACTT 33250

Search completed: June 20, 2004, 11:42:53
Job time : 564.069 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 114.921 Seconds
(without alignments)
739.327 Million cell updates/sec

Title: US-10-624-714-13

Perfect score: 20

Sequence: 1 gcgcgcaccacctcgactt 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : N_Geneseq_29Jan04:*

1: Geneseq_1980s:*

2: Geneseq_1990s:*

3: Geneseq_2000s:*

4: Geneseq_2001as:*

5: Geneseq_2001bs:*

6: Geneseq_2002as:*

7: Geneseq_2003as:*

8: Geneseq_2003bs:*

9: Geneseq_2003cs:*

10: Geneseq_2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20	100.0	867	7	ACA38314
C 2	20	100.0	870	7	ACA40752
C 3	20	100.0	110000	4	AAI99682_32
C 4	20	100.0	110000	4	AAI99682_32
C 5	18.4	92.0	849	7	ACA39852
C 6	18.4	92.0	849	7	ACA38167
C 7	17	85.0	162	2	AAI18071
C 8	17	85.0	176	2	AAI18072
C 9	17	85.0	495	7	ACA44066
C 10	16.8	84.0	858	5	AAH57237
C 11	16.8	84.0	858	7	ACA00707
C 12	16.8	84.0	1089	6	ABQ91626
C 13	16.8	84.0	1154	4	ABL20151
C 14	16.8	84.0	1204	4	ABL18389
C 15	16.8	84.0	1207	4	ABL20153
C 16	16.8	84.0	1404	4	ABL20149
C 17	16.8	84.0	2310	4	ABL05611
C 18	16.8	84.0	3250	4	ABL44062
C 19	16.8	84.0	3489	4	ABL18389
C 20	16.8	84.0	5110	4	ABL06782
C 21	16.8	84.0	6190	4	ABL20148
C 22	16.8	84.0	7108	4	ABL05610
C 23	16.8	84.0	7393	4	ABL20150

C 24	16.8	84.0	9160	4	ABL20152
C 25	16.8	84.0	349980	5	AAH68531
C 26	16.4	82.0	3326	5	AAF74625
C 27	16.4	82.0	3587	7	ADA83747
C 28	16.4	82.0	3600	4	ADL12302
C 29	16.4	82.0	3600	6	ABX30501
C 30	16.4	82.0	3600	6	AAI45544
C 31	16.4	82.0	3600	6	AAI45543
C 32	16.4	82.0	3600	6	AAI45542
C 33	16.4	82.0	3600	6	AAI45545
C 34	16.4	82.0	3600	9	ADD19025
C 35	16.4	82.0	8522	4	AAI27843
C 36	16.4	82.0	8522	9	ADB94646
C 37	16.4	82.0	14486	4	AAI27844
C 38	16.4	82.0	14486	9	ADB94647
C 39	16.2	81.0	677	3	AAI45084
C 40	16	80.0	293	6	ABL73207
C 41	16	80.0	416	3	AAI38335
C 42	16	80.0	1617	7	ABT23507
C 43	15.8	79.0	446	8	ACH42306
C 44	15.8	79.0	875	4	AAI59563
C 45	15.8	79.0	894	7	ABT56529
C 46	15.8	79.0	1117	4	AAI20402
C 47	15.8	79.0	1449	4	ABL25167
C 48	15.8	79.0	1641	2	AAI210823
C 49	15.8	79.0	2250	9	ABT63564
C 50	15.8	79.0	2706	7	ADA53073
C 51	15.8	79.0	2837	6	ABK35667
C 52	15.8	79.0	3449	4	ABL25166
C 53	15.8	79.0	4069	4	AAH17772
C 54	15.8	79.0	5236	4	AAH15656
C 55	15.8	79.0	5530	4	AAK52640
C 56	15.8	79.0	8144	7	AAI46634
C 57	15.8	79.0	10519	6	ABT10071
C 58	15.8	79.0	10519	6	ABT565009
C 59	15.8	79.0	11523	4	AAI45467
C 60	15.8	79.0	11523	6	ABK34641
C 61	15.8	79.0	11523	6	ABL33922
C 62	15.8	79.0	11523	6	ABK34001
C 63	15.8	79.0	11523	6	ABK28523
C 64	15.8	79.0	11523	7	ADA20382
C 65	15.8	79.0	11523	7	ADAB4189
C 66	15.8	79.0	11523	9	ADB54093
C 67	15.4	77.0	114	3	AAI49078
C 68	15.4	77.0	151	6	ABN25553
C 69	15.4	77.0	192	9	AUC87668
C 70	15.4	77.0	272	3	AAA49089
C 71	15.4	77.0	360	6	ABZ12519
C 72	15.4	77.0	504	6	ABQ38061
C 73	15.4	77.0	504	6	ABQ38060
C 74	15.4	77.0	547	3	AAI40683
C 75	15.4	77.0	596	3	AAI45054
C 76	15.4	77.0	598	3	AAI34034
C 77	15.4	77.0	625	6	ABQ42142
C 78	15.4	77.0	625	6	ABQ42143
C 79	15.4	77.0	633	4	ABL27051
C 80	15.4	77.0	861	7	ACA24986
C 81	15.4	77.0	935	3	AAI46244
C 82	15.4	77.0	939	6	ABK54079
C 83	15.4	77.0	939	6	ABK54083
C 84	15.4	77.0	978	3	AAA39924
C 85	15.4	77.0	1038	3	AAA49097
C 86	15.4	77.0	1050	6	ABQ90321
C 87	15.4	77.0	1085	3	AAA62229
C 88	15.4	77.0	1107	7	ABZ39173
C 89	15.4	77.0	1344	7	ACA39628
C 90	15.4	77.0	1401	4	AAI54457
C 91	15.4	77.0	1401	7	ACA19506
C 92	15.4	77.0	1588	7	ACA26952
C 93	15.4	77.0	1779	5	AAI70967
C 94	15.4	77.0	2469	3	AAA49095
C 95	15.4	77.0	2652	4	AAH26468
C 96	15.4	77.0	2652	4	AAH26470

97	15.4	77.0	3172	4	ABL27050	ABL27050 Drosophil	c 170	15.2	76.0	3042	6	ABX09145	ABX09145 Mycobact
98	15.4	77.0	4008	9	ACE07105	Novel cod	c 171	15.2	76.0	76363	9	ACEF30938	ACEF30938 Rice cult
99	15.4	77.0	4014	7	ACC79241	Human AVA	c 172	15.2	76.0	86765	9	ADD14752	ADD14752 Human src
100	15.4	77.0	4645	4	ABL14607	Drosophil	c 173	15.2	76.0	110000	2	AAV30458_0	AAV30458 Rhizobium
101	15.4	77.0	5615	6	ABN5962	Novel hum	c 174	15.2	76.0	110000	2	AAV30459_0	AAV30459 Rhizobium
102	15.4	77.0	5967	7	ACC79242	Human AVA	c 175	15.2	76.0	110000	4	AAI99682_28	Continuation (29 o
103	15.4	77.0	6233	9	ADC87652	Human mam	c 176	15.2	76.0	110000	4	AAI99683_28	Continuation (29 o
104	15.4	77.0	7407	9	ADB78685	Human pot	c 177	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool
105	15.4	77.0	7407	9	ADB78688	Human pot	c 178	15.2	76.0	135838	7	ABX34289	ABX34289 S. atrool
106	15.4	77.0	7407	9	ADB78686	Human pot	c 179	15.2	76.0	135838	7	ACA26539	ACA26539 Prokaryot
107	15.4	77.0	7407	9	ADB78683	Human pot	c 180	15.2	76.0	135838	7	ACA26539	ACA26539 Prokaryot
108	15.4	77.0	7407	9	ADB78687	Human pot	c 181	15.2	76.0	135838	7	ACA26539	ACA26539 Prokaryot
109	15.4	77.0	7411	9	ADD29557	Human tum	c 182	15.2	76.0	135838	7	ACA26539	ACA26539 Prokaryot
110	15.4	77.0	7413	5	AA874832	DNA encod	c 183	15.2	76.0	135838	7	ACA26539	ACA26539 Prokaryot
111	15.4	77.0	8059	3	AAA81747	N. mening	c 184	15.2	76.0	135838	7	ACA26539	ACA26539 Prokaryot
112	15.4	77.0	8309	4	AAI59190	Human pol	c 185	14.8	74.0	30	2	AAI59190	AAI59190 Asci-Hind
113	15.4	77.0	8309	4	AAI59190	Human pol	c 186	14.8	74.0	30	2	AAI59190	AAI59190 Asci-Hind
114	15.4	77.0	8309	4	AAI59190	Human pol	c 187	14.8	74.0	30	2	AAI59190	AAI59190 Asci-Hind
115	15.4	77.0	20348	4	ABL14606	Drosophil	c 188	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
116	15.4	77.0	20348	4	ABL14606	Drosophil	c 189	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
117	15.4	77.0	26970	9	ADC87695	Human mam	c 190	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
118	15.4	77.0	63164	3	AA633348	Streptomy	c 191	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
119	15.4	77.0	82993	6	ABX09140	Mycobacte	c 192	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
120	15.4	77.0	110000	4	AAI99682_12	Continuation (13 o	c 193	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
121	15.4	77.0	110000	4	AAI99682_12	Continuation (13 o	c 194	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
122	15.4	77.0	110000	4	AAI99682_12	Continuation (13 o	c 195	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
123	15.4	77.0	110000	4	AAI99682_12	Continuation (13 o	c 196	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
124	15.4	77.0	111393	7	AAI99682_12	Continuation (13 o	c 197	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
125	15.4	77.0	349980	3	AAI99682_12	Continuation (13 o	c 198	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
126	15.2	76.0	354	6	ABX09140	Mycobacte	c 199	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
127	15.2	76.0	354	6	ABX09140	Mycobacte	c 200	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
128	15.2	76.0	595	3	AAI99682_12	Continuation (13 o	c 201	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
129	15.2	76.0	654	3	AAI99682_12	Continuation (13 o	c 202	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
130	15.2	76.0	735	6	AAI99682_12	Continuation (13 o	c 203	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
131	15.2	76.0	971	6	AAI99682_12	Continuation (13 o	c 204	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
132	15.2	76.0	1086	7	AAI99682_12	Continuation (13 o	c 205	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
133	15.2	76.0	1221	9	ADC36217	Weed cont	c 206	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
134	15.2	76.0	1347	7	ABX56047	M. echino	c 207	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
135	15.2	76.0	1374	9	ABX56047	M. echino	c 208	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
136	15.2	76.0	1374	9	ABX56047	M. echino	c 209	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
137	15.2	76.0	1374	9	ABX56047	M. echino	c 210	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
138	15.2	76.0	1451	9	ADC36218	Weed cont	c 211	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
139	15.2	76.0	1590	7	AAI99682_12	Continuation (13 o	c 212	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
140	15.2	76.0	1843	4	ABL10155	Drosophil	c 213	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
141	15.2	76.0	1879	2	AAI99682_12	Continuation (13 o	c 214	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
142	15.2	76.0	1959	7	AAI99682_12	Continuation (13 o	c 215	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
143	15.2	76.0	2208	4	ABL23887	Drosophil	c 216	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
144	15.2	76.0	2244	9	ABX56047	M. echino	c 217	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
145	15.2	76.0	2289	7	AAI99682_12	Continuation (13 o	c 218	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
146	15.2	76.0	2649	9	ABX56047	M. echino	c 219	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
147	15.2	76.0	2694	7	AAI99682_12	Continuation (13 o	c 220	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
148	15.2	76.0	2887	3	AAI99682_12	Continuation (13 o	c 221	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
149	15.2	76.0	2932	4	AD16103	Human int	c 222	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
150	15.2	76.0	2932	4	AD16103	Human int	c 223	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
151	15.2	76.0	2933	2	AAV59665	Human sec	c 224	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
152	15.2	76.0	2933	6	ABX57385	Human cdn	c 225	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
153	15.2	76.0	3195	7	ABX57385	Human cdn	c 226	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
154	15.2	76.0	3276	6	ABX57385	Human cdn	c 227	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
155	15.2	76.0	3276	6	ABX57385	Human cdn	c 228	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
156	15.2	76.0	3276	6	ABX57385	Human cdn	c 229	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
157	15.2	76.0	3472	6	ABX57385	Human cdn	c 230	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
158	15.2	76.0	3776	6	ABX57385	Human cdn	c 231	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
159	15.2	76.0	4307	4	ABL10154	Drosophil	c 232	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
160	15.2	76.0	4649	9	ABX569161	C. neofo	c 233	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
161	15.2	76.0	7812	3	ACA64735	Mycobacte	c 234	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
162	15.2	76.0	8391	7	ACA64735	Mycobacte	c 235	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
163	15.2	76.0	9098	5	AAI99682_12	Continuation (13 o	c 236	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
164	15.2	76.0	9210	4	AAI99682_12	Continuation (13 o	c 237	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
165	15.2	76.0	9390	7	AAI99682_12	Continuation (13 o	c 238	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
166	15.2	76.0	16605	4	AAI99682_12	Continuation (13 o	c 239	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
167	15.2	76.0	16605	4	AAI99682_12	Continuation (13 o	c 240	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
168	15.2	76.0	18335	4	AAI99682_12	Continuation (13 o	c 241	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind
169	15.2	76.0	18335	7	AAI99682_12	Continuation (13 o	c 242	14.8	74.0	41	3	AAI59190	AAI59190 Asci-Hind

C 243	14.8	74.0	2175	4	ABL97752
C 244	14.8	74.0	2247	7	ACA37588
C 245	14.8	74.0	2463	7	ACA35940
C 246	14.8	74.0	2466	7	ACA3983
C 247	14.8	74.0	2511	7	ACA39804
C 248	14.8	74.0	2761	4	ABL24718
C 249	14.8	74.0	3083	7	ADA53528
C 250	14.8	74.0	3332	9	ADD44952
C 251	14.8	74.0	3401	4	ABL20460
C 252	14.8	74.0	3635	6	RAL45541
C 253	14.8	74.0	3668	9	ADD24893
C 254	14.8	74.0	3789	2	AAT63514
C 255	14.8	74.0	3882	7	ABZ37537
C 256	14.8	74.0	4873	4	AAC91400
C 257	14.8	74.0	5110	5	AAS29201
C 258	14.8	74.0	5110	5	ABA15205
C 259	14.8	74.0	5110	6	ABS68341
C 260	14.8	74.0	5110	9	ADC25463
C 261	14.8	74.0	5489	6	ABL34121
C 262	14.8	74.0	5726	4	ABL02959
C 263	14.8	74.0	6615	4	ABL26203
C 264	14.8	74.0	8343	4	ABL02958
C 265	14.8	74.0	9970	4	ABL35912
C 266	14.8	74.0	9970	7	ABX58900
C 267	14.8	74.0	10565	4	ABL26202
C 268	14.8	74.0	11905	7	AAI61225
C 269	14.8	74.0	13500	3	AZ45562
C 270	14.8	74.0	20922	9	ADC26981
C 271	14.8	74.0	25085	7	AAD54235
C 272	14.8	74.0	33529	5	AAS17367
C 273	14.8	74.0	35026	3	AAA64890
C 274	14.8	74.0	36033	9	ADB74375
C 275	14.8	74.0	50543	7	AAD54230
C 276	14.8	74.0	53905	7	ACF30939
C 277	14.8	74.0	59816	7	ABZ37516
C 278	14.8	74.0	59816	7	ABZ37515
C 279	14.8	74.0	67251	9	ADC25995
C 280	14.8	74.0	68750	3	AAS5887
C 281	14.8	74.0	71989	3	AAZ29349
C 282	14.8	74.0	110000	3	AAF22305
C 283	14.8	74.0	110000	4	AAI99682
C 284	14.8	74.0	110000	4	AAI99683
C 285	14.8	74.0	110000	4	AAI99682
C 286	14.8	74.0	110000	4	AAI99683
C 287	14.8	74.0	110000	4	AAI99683
C 288	14.8	74.0	110000	4	AAI99683
C 289	14.8	74.0	110000	4	AAI99683
C 290	14.8	74.0	110000	4	AAI99683
C 291	14.8	74.0	125401	4	AAI99683
C 292	14.8	74.0	125401	4	AAI99683
C 293	14.8	74.0	207433	5	ABZ72040
C 294	14.8	74.0	207433	3	ABZ72040
C 295	14.4	72.0	45	2	AAQ63170
C 296	14.4	72.0	45	2	AAQ63170
C 297	14.4	72.0	161	3	AAC18228
C 298	14.4	72.0	259	6	ABQ55809
C 299	14.4	72.0	307	6	ABZ27471
C 300	14.4	72.0	308	7	ABX43590

ALIGNMENTS

RESULT 1
 ID ACA38314/C
 AC ACA38314; standard; DNA; 867 BP.
 XX ACA38314;
 XX ACA38314;
 DT 19-JUN-2003 (first entry)
 DE Prokaryotic essential gene #19971.
 XX

Query Match 100.0%; Score 20; DB 7; Length 867;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCCGACACCTCGACTT 20
 DB 36 GCGCCGACACCTCGACTT 17

KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX Mycobacterium bovis.
 OS WO200277193-A2.
 PN 03-OCT-2002.
 PD 21-MAR-2002; 2002WO-US009107.
 PF 21-MAR-2001; 2001US-00815242.
 XX 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI: 2003-029926/02.
 DR P-PSDB; ABU34444.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 14; SEQ ID NO 26184; 1766pp; English.
 XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 867 BP; 136 A; 269 C; 314 G; 148 T; 0 U; 0 Other;

RESULT 2

ACA40752/c

ID ACA40752 standard; DNA; 870 BP.

XX AC

XX ACA40752;

XX DT

XX 19-JUN-2003 (first entry)

XX DE

XX Prokaryotic essential gene #22409.

XX KW

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KN

XX drug design; gene.

XX OS

XX Mycobacterium tuberculosis.

XX PN

XX WO200277183-A2.

XX PD

XX 03-OCT-2002.

XX PF

XX 21-MAR-2002; 2002WO-US009107.

XX PR

XX 21-MAR-2001; 2001US-00815242.

XX PR

XX 06-SEP-2001; 2001US-00948993.

XX PR

XX 28-OCT-2001; 2001US-0342923P.

XX PR

XX 08-FEB-2002; 2002US-00072851.

XX PR

XX 06-MAR-2002; 2002US-0362699P.

XX PA

XX (ELIT-) ELITRA PHARM INC.

XX XX

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX XX

XX WPI: 2003-029926/02.

XX DR

XX P-PSDB; ABU36882.

XX XX

XX New antisense nucleic acids, useful for identifying proteins or screening

XX PT

XX for homologous nucleic acids required for cellular proliferation to

XX PT

XX isolate candidate molecules for rational drug discovery programs.

XX XX

XX Claim 14; SEQ ID NO 28622; 1766pp; English.

XX PS

XX The invention relates to an isolated nucleic acid comprising any one of

XX CC

XX the 6213 antisense sequences given in the specification where expression

XX CC

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC

XX encoding a polypeptide whose expression is inhibited by the antisense

XX CC

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC

XX polypeptide or its fragment whose expression is inhibited by the

XX CC

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC

XX proliferation or the activity of a gene in an operon required for

XX CC

XX proliferation; (7) identifying a compound that influences the activity of

XX CC

XX the gene product or that has an activity against a biological pathway

XX CC

XX required for proliferation, or that inhibits cellular proliferation; (8)

XX CC

XX identifying a gene required for cellular proliferation or the biological

XX CC

XX pathway in which a proliferation-required gene or its gene product lies

XX CC

XX on a gene on which the test compound that inhibits proliferation of an

XX CC

XX organism's activity; (9) manufacturing an antibiotic; (10) profiling a

XX CC

XX product is overexpressed or underexpressed; (12) determining the extent

XX CC

XX to which each of the strains is present in a culture or collection of

XX CC

XX strains; or (13) identifying the target of a compound that inhibits the

XX CC

XX proliferation of an organism. The antisense nucleic acids are useful for

XX CC

XX identifying proteins or screening for homologous nucleic acids required

XX CC

XX for cellular proliferation to isolate candidate molecules for rational

XX CC

XX drug discovery programs, or for screening homologous nucleic acids

XX CC

XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

XX CC

XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

XX CC

XX prokaryotic essential genes. Note: the sequence data for this patent did

XX CC

XX not form part of the printed specification, but was obtained in

XX CC

XX electronic format directly from WIPO at

XX CC

XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 870 BP; 137 A; 269 C; 315 G; 149 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 7; Length 870;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20

Db 36 GGGCGGCACCACTCGACTT 17

RESULT 3

AAI99682_12

Continuation (33 of 45) of AAI99682 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP Fragment Name Begin End

WP AAI99682_00 1 110000

WP AAI99682_01 100001 210000

WP AAI99682_02 200001 310000

WP AAI99682_03 300001 410000

WP AAI99682_04 400001 510000

WP AAI99682_05 500001 610000

WP AAI99682_06 600001 710000

WP AAI99682_07 700001 810000

WP AAI99682_08 800001 910000

WP AAI99682_09 900001 1010000

WP AAI99682_10 1000001 1110000

WP AAI99682_11 1100001 1210000

WP AAI99682_12 1200001 1310000

WP AAI99682_13 1300001 1410000

WP AAI99682_14 1400001 1510000

WP AAI99682_15 1500001 1610000

WP AAI99682_16 1600001 1710000

WP AAI99682_17 1700001 1810000

WP AAI99682_18 1800001 1910000

WP AAI99682_19 1900001 2010000

WP AAI99682_20 2000001 2110000

WP AAI99682_21 2100001 2210000

WP AAI99682_22 2200001 2310000

WP AAI99682_23 2300001 2410000

WP AAI99682_24 2400001 2510000

WP AAI99682_25 2500001 2610000

WP AAI99682_26 2600001 2710000

WP AAI99682_27 2700001 2810000

WP AAI99682_28 2800001 2910000

WP AAI99682_29 2900001 3010000

WP AAI99682_30 3000001 3110000

WP AAI99682_31 3100001 3210000

WP AAI99682_32 3200001 3310000

WP AAI99682_33 3300001 3410000

WP AAI99682_34 3400001 3510000

WP AAI99682_35 3500001 3610000

WP AAI99682_36 3600001 3710000

WP AAI99682_37 3700001 3810000

WP AAI99682_38 3800001 3910000

WP AAI99682_39 3900001 4010000

WP AAI99682_40 4000001 4110000

WP AAI99682_41 4100001 4210000

WP AAI99682_42 4200001 4310000

WP AAI99682_43 4300001 4410000

WP AAI99682_44 4400001 4411529

Query Match 100.0%; Score 20; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCGGCACCACTCGACTT 20

Db 39432 GGGCGGCACCACTCGACTT 39451

RESULT 4

AAI99683_32
Continuation (33 of 44) of AAI99683 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Query Match 100.0%; Score 20; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GCGCGCACCCTCGACTT 20
Db 33708 GCGCGCACCCTCGACTT 33727

RESULT 5
ACA39852/c
ID ACA39852 standard; DNA; 849 BP.
XX AC ACA39852;
XX DT 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #21509.
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX KW drug design; gene.
XX OS Mycobacterium leprae.

PN WO200277183-A2.
XX 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX DR WPI; 2003-029926/02.
XX DR F-PSDB; ABU35982.
XX PT New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 14; SEQ ID NO 27722; 17669p; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway; (8)
XX CC required for proliferation, or that inhibits cellular proliferation; (9)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than S. aureus, S. typhimurium,
XX CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX CC prokaryotic essential genes. Note: The sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 849 BP; 150 A; 236 C; 279 G; 184 T; 0 U; 0 Other;

Query Match 92.0%; Score 18.4; DB 7; Length 849;
Best Local Similarity 95.0%; Pred. No. 98;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GCGCGCACCCTCGACTT 20
Db 36 GCGCGCACCCTCGACTT 17

RESULT 6
ACA38167/c
ID ACA38167 standard; DNA; 849 BP.
XX OS Mycobacterium leprae.

AC ACA38167;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #19824.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
OS Mycobacterium avium.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342922P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
PR P-PSDB; ABU34297.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 26037; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 849 BP; 128 A; 301 C; 297 G; 121 T; 0 U; 2 Other;
Query Match 92.0%; Score 18.4; DB 7; Length 849;
Best Local Similarity 95.0%; Pred. No. 98;
Query Match 85.0%; Score 17; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCGCGCGACCACTGACTT 20
Db 36 GCGCGCGACCACTGACTT 17
RESULT 7
AAAX18071/c
ID AAX18071 standard; DNA; 162 BP.
XX
AC AAX18071;
XX
DT 04-MAY-1999 (first entry)
XX
XX Coding sequence for human D2H binding protein DAB3.
XX
XX Gastro-intestinal transport receptor; binding protein; hsr; HPT1; D2H;
KW hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW therapeutic agent delivery; therapy; ss.
XX
XX Homo sapiens.
XX
XX WO9851325-A2.
XX
XX 19-NOV-1998.
XX
XX 15-MAY-1998; 98WO-US010088.
XX
XX 15-MAY-1997; 97US-0046595P.
XX
XX (CYTO-) CYTOGEN CORP.
XX (ELAN-) ELAN CORP PLC.
XX
XX Alvarez VL, Omahony DJ, Lambkin IJ, Patterson CA, Singleton J;
PI Belinka BA, Carter JM, Cagney GW;
XX
XX WPI; 1999-009568/01.
XX
XX New proteins that bind specifically to receptors in the gastro-intestinal
PT tract and related nucleic acid-chimaeras and antibodies, used to
PT deliver therapeutic or diagnostic agents to, or through, the
XX gastrointestinal tract, e.g. insulin or leuprolide.
XX
XX Claim 49; Page 57; 294pp; English.
XX
XX This sequence encodes a peptide that specifically binds to the human D2H
CC protein. The invention relates to purified proteins (I) that bind
CC specifically to at least one of the gastro-intestinal (GI) tract
CC receptors human intestinal peptide-associated transporter (HPT1), hPEPT1,
CC D2H and human sucrose-isomaltase complex (hsr). (I) provide active
CC transport of therapeutic agents through human and animal GI tissue (into
CC the blood) for in vivo delivery, particularly for treatment or prevention
CC of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC migraine, or angina pectoris. Specifically they are used to deliver
CC insulin or leuprolide, but many other suitable therapeutic agents are
CC disclosed, including genes or inhibitory nucleic acid, imaging agents and
CC antigens. (I) may also provide targeting to the GI tract. Other uses of
CC (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunoassays. The antibodies of (I) are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
XX etc., also for peptide purification and immobilisation
XX
SQ Sequence 162 BP; 30 A; 58 C; 42 G; 32 T; 0 U; 0 Other;
Query Match 85.0%; Score 17; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGA 17
|||||
Db 26 GCGCGCACCACTCGA 10

RESULT 8
AA18072/c
ID AAX18072 standard; DNA; 176 BP.
XX AC
XX AC AAX18072;
XX
XX
DT 04-MAY-1999 (first entry)
XX
XX
DE Coding sequence for human D2H binding protein DAB7.
XX
XX Gastro-intestinal transport receptor; binding protein; hSI; HPT1; D2H;
XX hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
XX intestinal peptide-associated transporter; hypertension; diabetes;
XX osteoporosis; haemophilia; anaemia; cancer; migraines; angina pectoris;
XX therapeutic agent delivery; therapy; ss.
XX
XX Homo sapiens.
XX
XX WO9851325-A2.
XX
XX 19-NOV-1998.
XX
XX 15-MAY-1998; 98WO-US010088.
XX
XX 15-MAY-1997; 97US-0046595P.
XX
XX (CYTO-) CYTOGEN CORP.
XX (ELAN-) ELAN CORP PLC.
XX
XX Alvarez VL, Omahony DJ, Lambkin J, Patterson CA, Singleton J;
XX Belinka BA, Carter JM, Cagney GM;
XX WPI; 1999-009568/01.
XX
XX New proteins that bind specifically to receptors in the gastro-intestinal
XX tract and related nucleic acid - chimera and antibodies, used to
XX deliver therapeutic or diagnostic agents to, or through, the
XX gastrointestinal tract, e.g. insulin or leuprolide.
XX
XX Claim 49; Page 57; 294pp; English.

CC This sequence encodes a peptide that specifically binds to the human D2H
XX protein. The invention relates to purified proteins (I) that bind
XX specifically to at least one of the gastro-intestinal (GI) tract
XX receptors human intestinal peptide-associated transporter (HPT1), hPEPT1,
XX D2H and human sucrose-isomaltase complex (hSI). (I) provide active
XX transport of therapeutic agents through human and animal GI tissue (into
XX the blood) for in vivo delivery, particularly for treatment or prevention
XX of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
XX migraine, or angina pectoris. Specifically they are used to deliver
XX insulin or leuprolide, but many other suitable therapeutic agents are
XX disclosed, including genes or inhibitory nucleic acid, imaging agents and
XX antigens. (I) may also provide targeting to the GI tract. Other uses of
XX (I) are: (i) to determine the level of specified receptors in a sample
XX (in a binding assay); and (ii) to screen for molecules that bind (I).
XX immunogenic analogues or derivatives of (I) are used to raise antibodies
XX and in immunoassays. The antibodies are used to locate, detect and
XX measure (I), e.g. for imaging, monitoring treatment, tissue analysis
XX etc., also for peptide purification and immobilisation

SQ Sequence 176 BP; 28 A; 46 C; 60 G; 42 T; 0 U; 0 Other;
Query Match 85.0%; Score 17; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGA 17
|||||

Db 26 GCGCGCACCACTCGA 10

RESULT 9
ACA4066/c
ID ACA4066 standard; DNA; 495 BP.
XX AC
XX AC ACA4066;
XX
XX 19-JUN-2003 (first entry)
XX
XX Prokaryotic essential gene #25723.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Pseudomonas putida.
XX OS
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX P-PSDB; ABU40196.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids, required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 31936; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX prokaryotic essential genes. Note: The sequence data for this patent did
XX not form part of the printed specification, but was obtained in

PN WO200255655-A2.
XX 18-JUL-2002.
XX 14-JAN-2002; 2002WO-N0000019.
XX 12-JAN-2001; 2001NO-00000235.
XX 12-JAN-2001; 2001NO-00000239.
XX (UNIF-) UNIFOB STIFTTELSEN UNIV BERGEN.
XX (TIGR-) TIGR.
XX Birkeland NK, Bidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX WPI; 2002-557818/59.
XX Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprises polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes.
XX Claim 14; Page 613; 678pp; English.
XX The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
CC invention
XX Sequence 1089 BP; 260 A; 321 C; 338 G; 170 T; 0 U; 0 Other;
SQ Query Match 84.0%; Score 16.8; DB 6; Length 1089;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20
Db 870 GCGCGCACCACTCGACTT 899
RESULT 13
ABL20151
ID ABL20151 standard; DNA; 1154 BP.
XX ABL20151;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 11926.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 11926; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1154 BP; 285 A; 310 C; 277 G; 282 T; 0 U; 0 Other;
SQ Query Match 84.0%; Score 16.8; DB 4; Length 1154;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20
Db 159 GCTCCGACCACTCGACTT 178
RESULT 14
ABL18389
ID ABL18389 standard; DNA; 1204 BP.
XX ABL18389;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 6640.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX Claim 1; SEQ ID NO 6640; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 11926; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1154 BP; 285 A; 310 C; 277 G; 282 T; 0 U; 0 Other;
SQ Query Match 84.0%; Score 16.8; DB 4; Length 1154;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20
Db 159 GCTCCGACCACTCGACTT 178
RESULT 14
ABL18389
ID ABL18389 standard; DNA; 1204 BP.
XX ABL18389;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 6640.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 6640; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the


```

XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB61508.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 11315; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2310 BP; 608 A; 754 C; 571 G; 377 T; 0 U; 0 Other;
XX
XX Query Match 84.0%; Score 16.8; DB 4; Length 2310;
XX Best Local Similarity 90.0%; Pred. No. 4.7e+02;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX Qy 1 GCGCGCACCCACTCGACTT 20
XX 1920 GCGCGCACCCACTCGACTT 1901
XX
XX RESULT 18
XX ABL14062/c
XX ID ABL14062 standard; cDNA; 3250 BP.
XX
XX AC ABL14062;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36658.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX
XX KW pharmaceutical; gene; ss.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US009231.
XX
XX PR 23-MAR-2000; 2000US-0191637P.
XX
XX PR 11-JUL-2000; 2000US-00614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB69959.

```

```

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 36668; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3250 BP; 975 A; 697 C; 755 G; 823 T; 0 U; 0 Other;
XX
XX Query Match 84.0%; Score 16.8; DB 4; Length 3250;
XX Best Local Similarity 90.0%; Pred. No. 4.7e+02;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX Qy 1 GCGCGCACCCACTCGACTT 20
XX 378 GCTCCGACCATCTCGACTT 359
XX
XX RESULT 19
XX ABL18388/c
XX ID ABL18388 standard; DNA; 3489 BP.
XX
XX AC ABL18388;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 6637.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX
XX KW pharmaceutical; gene; ds.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US009231.
XX
XX PR 23-MAR-2000; 2000US-0191637P.
XX
XX PR 11-JUL-2000; 2000US-00614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 6637; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3250 BP; 975 A; 697 C; 755 G; 823 T; 0 U; 0 Other;

```

CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 3489 BP; 1067 A; 715 C; 723 G; 984 T; 0 U; 0 Other;
 Query Match 84.0%; Score 16.8; DB 4; Length 3489;
 Best Local Similarity 90.0%; Pred. No. 4.7e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 Db 2062 GCTCGCACCACTCGACTT 2043

RESULT 20
 ABL06782/c
 ID ABL06782 standard; cDNA; 5110 BP.
 XX
 AC ABL06782;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 14828.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX

PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX

PS Claim 1; SEQ ID NO 14828; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABBS7072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 5110 BP; 1321 A; 1226 C; 1260 G; 1303 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 4; Length 5110;
 Best Local Similarity 90.0%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 Db 302 GCGCGCACCACTCGACTT 283

RESULT 21
 ABL20148/c
 ID ABL20148 standard; DNA; 6190 BP.
 XX
 AC ABL20148;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11917.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 11917; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
 CC ABBS7072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6190 BP; 1866 A; 1245 C; 1340 G; 1739 T; 0 U; 0 Other;

Query Match 84.0%; Score 16.8; DB 4; Length 6190;
 Best Local Similarity 90.0%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 Db 2062 GCTCGCACCACTCGACTT 2043

RESULT 22
 ABL05610/c
 ID ABL05610 standard; cDNA; 7108 BP.
 XX
 AC ABL05610;
 XX

DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11312.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX

PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR P-FSDB; ABB61507.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 11312; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 7108 BP; 1909 A; 1824 C; 1701 G; 1674 T; 0 U; 0 Other;
 SQ
 Query Match 84.0%; Score 16.8; DB 4; Length 7108;
 Best Local Similarity 90.0%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCGCACCACTCGACTT 20
 DB 5718 GCGCGCGCACCACTCGACTT 5699
 RESULT 23
 ABL20150/c
 ID ABL20150 standard; DNA; 7393 BP.
 XX
 AC ABL20150;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11923.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 11929; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 7108 BP; 1909 A; 1824 C; 1701 G; 1674 T; 0 U; 0 Other;
 SQ

XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 11923; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 7393 BP; 2294 A; 1463 C; 1583 G; 2053 T; 0 U; 0 Other;
 SQ
 Query Match 84.0%; Score 16.8; DB 4; Length 7393;
 Best Local Similarity 90.0%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCGCACCACTCGACTT 20
 DB 2062 GCTCGCGCACCACTCGACTT 2043
 RESULT 24
 ABL20152/c
 ID ABL20152 standard; DNA; 9160 BP.
 XX
 AC ABL20152;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11929.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 XX WPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 XX Claim 1; SEQ ID NO 11929; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the

XX DE Human GLI gene.
XX KW human; marker; expressed sequence tag; EST; arabidopsis; tumour;
XX KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;
XX KW vaccine; ds; gene.
XX OS Homo sapiens.
XX PN WO2002103028-A2.
XX PD 27-DEC-2002.
XX PF 30-MAY-2002; 2002WO-IB004189.
XX PR 30-MAY-2001; 2001US-0293999P.
XX PR 22-OCT-2001; 2001US-0330457P.
XX PR 19-FEB-2002; 2002US-0357144P.
XX PA (BIOM-) BIOMEDICAL CENT.
XX PI Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LI;
XX DR WPI: 2003-175241/17.
XX DR P-PSDB; ADA83748.
XX PT Determining if a nucleic acid is a marker for a phenotype/cell type of
XX PT interest, by global comparison of expressed sequence tags known to be
XX PT expressed in the phenotype/cell type with all ESTs expressed in normal
XX PT tissue.
XX PS Claim 23; Page 104-106; 516pp; English.
XX CC The invention relates to a novel method for determining if a nucleic acid
CC is a marker for a predetermined phenotype/cell type of interest from a
CC biological species. The method comprises performing a global comparison
CC of a group of expressed sequence tags (ESTs) known to be expressed in the
CC phenotype/cell type of interest with all ESTs expressed in normal tissue
CC in order to identify ESTs that are preferentially expressed in the
CC phenotype/cell of interest. A method of the invention is useful for
CC determining whether a nucleic acid is a marker for a predetermined
CC phenotype or cell type of interest from a biological species, preferably
CC Arabidopsis or human. The cell type of interest is an abnormal cell such
CC as a tumour cell, and the predetermined phenotype is a stress-induced
CC phenotype such as hyperosmotic stress or high salt conditions. A method
CC of the invention is also useful for determining the progression of colon
CC cancer in a human, for detecting a tumour cell, and for regulating or
CC preventing the growth of a tumour cell. An antibody of the invention is
CC useful for detecting the absence or presence of peptides encoded by
CC tumour-associated markers. A polypeptide of the invention is useful as an
CC immunogen for vaccinating an animal. The present sequence encodes a
CC tumour-associated antigen of the invention.
XX SQ Sequence 3587 BP; 798 A; 1143 C; 944 G; 702 T; 0 U; 0 Other;
Query Match 82.0%; Score 16.4; DB 7; Length 3587;
Best Local Similarity 94.4%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCCGCACACCTCGACTT 20
DB 3003 GCAGCACCACCTCGACTT 3020
RESULT 28
RAD12302
ID AAD12302 standard; CDNA; 3600 BP.
XX AC AAD12302;
XX XX
XX DT 16-OCT-2001 (first entry)
XX DE Human Cubitus interruptus (Ci) homologue, GLI-1 cDNA.
XX DE Human; glioma-associated oncogene-1 associated disease; infection;
XX KW inflammation; tumour formation; cytostatic; antiinflammatory; gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 79..3399
XX FT CDS /*tag= a
XX FT /*product= "Glioma-associated oncogene-1 protein"

XX KW Human; transgenic non-human animal; Cubitus interruptus; Ci; GLI-1;
XX KW basal cell carcinoma; BCC model system; tumour; screening; anti-cancer;
XX KW trichioepithelioma; cylindroma; trichoblastoma; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 79..3399
XX FT CDS /*tag= a
XX FT /*product= "Human Ci homologue, GLI-1"
XX PN WO200156376-A1.
XX PD 09-AUG-2001.
XX PF 02-FEB-2001; 2001WO-SE000204.
XX PR 03-FEB-2000; 2000SE-00000345.
XX PA (KARO-) KAROLINSKA INNOVATIONS AB.
XX PI Toftgard R;
XX DR WPI: 2001-488828/53.
XX DR P-PSDB; AA306644.
XX PT Transgenic non-human animal useful as basal cell carcinoma model system
XX PT to identify anti-cancer drug candidates, overexpresses transgene encoding
XX PT GLI-1 protein which is a human homologue to Cubitus interruptus.
XX PS Claim 6; Page 25-26; 33pp; English.
XX CC The present invention relates to a transgenic non-human animal comprising
CC a transgene containing a nucleic acid encoding a human Cubitus
CC interruptus (Ci) homologue protein, GLI-1. The transgenic non-human
CC animal is useful as basal cell carcinoma (BCC) model system since it
CC overexpresses GLI-1 which leads to development of tumours resembling
CC human BCC. Thus it is also useful for screening anti-cancer drug
CC candidates and evaluating whether it affects BCC, trichioepitheliomas,
CC cylindromas and trichoblastomas. The present sequence is a cDNA encoding
XX SQ Sequence 3600 BP; 785 A; 1151 C; 949 G; 705 T; 0 U; 0 Other;
Query Match 82.0%; Score 16.4; DB 4; Length 3600;
Best Local Similarity 94.4%; Pred. No. 6.9e-02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCCGCACACCTCGACTT 20
DB 3034 GCAGCACCACCTCGACTT 3051
RESULT 29
ABK30501
ID ABK30501 standard; DNA; 3600 BP.
XX AC ABK30501;
XX XX
XX DT 23-APR-2002 (first entry)
XX DE Human glioma-associated oncogene-1 DNA sequence.
XX KW Human; glioma-associated oncogene-1 associated disease; infection;
XX KW inflammation; tumour formation; cytostatic; antiinflammatory; gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT 79..3399
XX FT CDS /*tag= a
XX FT /*product= "Glioma-associated oncogene-1 protein"

XX PN US6329203-B1.
XX PD 11-DEC-2001.
XX PF 08-SEP-2000; 2000US-00657042.
XX PR 08-SEP-2000; 2000US-00657042.
XX PA (ISIS-) ISIS PHARM INC.
XX PI Bennett CF, Wyatt J;
XX WPI; 2002-138363/18.
XX P-PSDB; AAU12085.
XX Novel antisense compounds targeted to nucleic acids encoding glioma-associated oncogene-1, for modulating the gene expression and treating diseases associated with expression of the oncogene in humans.
XX Example 13; Col 47-56; 43pp; English.
XX The present invention relates to antisense compounds and methods for modulating the expression of human glioma-associated oncogene-1. The antisense compounds, particularly antisense oligonucleotides (ABK30509-ABK30586), target and inhibit the expression of human glioma-associated oncogene-1. The antisense compounds are useful for inhibiting the expression of human glioma-associated oncogene-1 in human cells or tissues and for treating an animal, particularly a human suspected of having or being prone to a disease or condition associated with expression of glioma-associated oncogene-1. The compounds are useful for diagnostics, therapeutics and as research reagent, e.g. prophylactically to prevent or delay infection, inflammation or tumour formation. The antisense compounds are safely and effectively administered to humans. The present sequence represents human glioma-associated oncogene-1 DNA
SQ Sequence 3600 BP; 785 A; 1161 C; 949 G; 705 T; 0 U; 0 Other;
Query Match 82.0%; Score 16.4; DB 6; Length 3600;
Best Local Similarity 94.4%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCCGACCACCTCGACTT 20
DB 3034 GCAGCACCACCTCGACTT 3051
RESULT 30
AAL45544
ID AAL45544 standard; cDNA; 3600 BP.
XX AC AAL45544;
XX DT 11-JUN-2002 (first entry)
XX DE Human Gli1 coding sequence SEQ ID NO: 16.
XX KW Gli1; screening method; bone induction; cartilage induction; orthopaedic disease; dental disease; osteoporosis; hyperosteoegenesis; osteopathic; antiarthritic; vulnary; immunosuppressive; human; hyperchondrogenesis; gene; ss.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT CDS 79..3399
XX FT /*tag= a
XX FT /product= "Gli1"
XX FT /transl_except= (pos:2875..2877,aa:Glu)
XX PN WO200211752-A1.
XX PD 14-FEB-2002.

XX 03-AUG-2001; 2001WO-JF006688.
XX PF 04-AUG-2000; 2000JP-00242767.
XX PR (TAKE) TAKEDA CHEM IND LTD.
XX PA Hikichi Y;
XX PI WPI; 2002-241709/29.
XX DR P-PSDB; AAO17111.
XX PT Promotion of bone and cartilage formation using Gli1 protein or DNA encoding it for treatment of skeletal disorders.
XX Claim 6; Page 115-117; 154pp; Japanese.
XX The present invention relates to agents for the promotion of bone and cartilage formation which contain as the active component a Gli1 protein or a DNA encoding a Gli1 protein. The agents can be used in the prevention, treatment and diagnosis of bone and cartilage disorders including bone fractures, joint deformation, osteoarthritis, osteoporosis, cartilage damage, trauma, bone formation defects, cartilage formation defects, bone defects, dental disease, hyperosteoegenesis and hyperchondrogenesis, and for use in cosmetic and therapeutic bone transplantation. The present sequence is a human Gli1 coding sequence described in the exemplification of the invention
SQ Sequence 3600 BP; 786 A; 1161 C; 948 G; 705 T; 0 U; 0 Other;
Query Match 82.0%; Score 16.4; DB 6; Length 3600;
Best Local Similarity 94.4%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCCGACCACCTCGACTT 20
DB 3034 GCAGCACCACCTCGACTT 3051
Search completed: June 20, 2004, 10:17:02
Job time : 133.046 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 934.603 Seconds
(without alignments)

639.034 Million cell updates/sec

Title: US-10-624-714-13

Perfect score: 20

Sequence: 1 gcgcgcaccaccctcgactt 20

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 300 summaries

Database :

EST:

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estom:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	85.0	1040	29	AG056593 Pan trogl
2	16.8	84.0	223	12	BI470080 saf38e11.
3	16.8	84.0	298	12	BI239271 RE35720.5
4	16.8	84.0	305	13	BQ102855 UUGC0247

5	16.8	84.0	309	12	BI212140
6	16.8	84.0	310	12	BI172928
7	16.8	84.0	333	10	BF486064
8	16.8	84.0	333	12	BI241833
9	16.8	84.0	335	12	BI363139
10	16.8	84.0	336	12	BI240825
11	16.8	84.0	345	12	BI578091
12	16.8	84.0	346	12	BI227580
13	16.8	84.0	356	12	BI361366
14	16.8	84.0	364	9	AI061909
15	16.8	84.0	378	9	AA201741
16	16.8	84.0	393	12	BI243125
17	16.8	84.0	397	12	BI369591
18	16.8	84.0	399	12	BI239432
19	16.8	84.0	399	12	BI355076
20	16.8	84.0	404	12	BI357752
21	16.8	84.0	414	9	AA695426
22	16.8	84.0	414	12	BI577942
23	16.8	84.0	418	13	BQ103130
24	16.8	84.0	419	13	BQ103552
25	16.8	84.0	427	12	BI163420
26	16.8	84.0	427	12	BI213953
27	16.8	84.0	427	12	BI375669
28	16.8	84.0	427	12	BI484865
29	16.8	84.0	428	12	BI365791
30	16.8	84.0	428	12	BI366732
31	16.8	84.0	429	12	BI239241
32	16.8	84.0	430	12	BI213411
33	16.8	84.0	441	12	BI165456
34	16.8	84.0	448	12	BI212897
35	16.8	84.0	452	12	BI233186
36	16.8	84.0	453	12	BI240875
37	16.8	84.0	454	9	AA201722
38	16.8	84.0	456	12	BI375564
39	16.8	84.0	457	12	BI212874
40	16.8	84.0	457	12	BI363271
41	16.8	84.0	458	12	BI481967
42	16.8	84.0	460	12	BI608622
43	16.8	84.0	461	12	BI241749
44	16.8	84.0	462	12	BI370712
45	16.8	84.0	465	12	BI164419
46	16.8	84.0	467	9	AI135940
47	16.8	84.0	467	9	AA201915
48	16.8	84.0	467	12	BI582929
49	16.8	84.0	468	12	BI486175
50	16.8	84.0	475	12	BI608870
51	16.8	84.0	476	12	BI167792
52	16.8	84.0	476	12	BI579620
53	16.8	84.0	476	12	BI581169
54	16.8	84.0	477	10	BF501960
55	16.8	84.0	480	12	BI172447
56	16.8	84.0	482	12	BI228973
57	16.8	84.0	483	10	BF419339
58	16.8	84.0	484	9	AI296025
59	16.8	84.0	486	12	BI356678
60	16.8	84.0	486	12	BI372713
61	16.8	84.0	487	12	BI355519
62	16.8	84.0	487	12	BI362624
63	16.8	84.0	489	9	AA695778
64	16.8	84.0	492	12	BI171646
65	16.8	84.0	492	12	BI365023
66	16.8	84.0	493	12	BI171638
67	16.8	84.0	493	12	BI605905
68	16.8	84.0	494	12	BI363780
69	16.8	84.0	501	12	BI593455
70	16.8	84.0	501	12	BI627342
71	16.8	84.0	501	13	BQ103380
72	16.8	84.0	503	12	BI584671
73	16.8	84.0	503	12	BI627199
74	16.8	84.0	504	29	TA226802Q
75	16.8	84.0	506	12	BI581210
76	16.8	84.0	507	12	BI608523
77	16.8	84.0	515	12	BI170929

C

78	16.8	84.0	516	12	BI213717	RE19185.5	151	16.8	84.0	590	9	AI519385	AI519385
79	16.8	84.0	516	12	BI214379	RE5438.5	152	16.8	84.0	590	10	BE585438	BE585438
80	16.8	84.0	516	12	BI368926	RE5438.5	153	16.8	84.0	591	9	AA438350	AA438350
81	16.8	84.0	516	29	CG233095	OGVCR89TV	154	16.8	84.0	591	12	BI574950	BI574950
82	16.8	84.0	518	12	BI613418	RE42567.5	155	16.8	84.0	591	12	BI589287	BI589287
83	16.8	84.0	519	12	BI618036	RE48552.5	156	16.8	84.0	592	9	AA390375	AA390375
84	16.8	84.0	520	12	BI611182	RE17337.5	157	16.8	84.0	592	9	AI519385	AI519385
85	16.8	84.0	525	9	AI546641	SD09804.5	158	16.8	84.0	593	12	BI364594	BI364594
86	16.8	84.0	526	12	BI368300	RE54249.5	159	16.8	84.0	593	12	BI364594	BI364594
87	16.8	84.0	527	12	BI368633	RE49794.5	160	16.8	84.0	594	12	BI615292	BI615292
88	16.8	84.0	528	12	BI580722	RE74945.5	161	16.8	84.0	594	12	BI615292	BI615292
89	16.8	84.0	528	9	AI294617	LP08032.5	162	16.8	84.0	595	12	BI629243	BI629243
90	16.8	84.0	528	12	BI681332	RE08688.5	163	16.8	84.0	596	12	BI213057	BI213057
91	16.8	84.0	528	12	BI170588	RE12104.5	164	16.8	84.0	596	12	BI482586	BI482586
92	16.8	84.0	528	12	BI170656	RE12181.5	165	16.8	84.0	596	12	BI627129	BI627129
93	16.8	84.0	528	12	BI173768	RE16967.5	166	16.8	84.0	597	12	BI241644	BI241644
94	16.8	84.0	528	12	BI228008	RE25417.5	167	16.8	84.0	597	12	BI371839	BI371839
95	16.8	84.0	528	12	BI229663	RE27490.5	168	16.8	84.0	599	12	BI365336	BI365336
96	16.8	84.0	528	12	BI234104	RE29745.5	169	16.8	84.0	599	12	BI366350	BI366350
97	16.8	84.0	528	12	BI235906	RE31860.5	170	16.8	84.0	599	12	BI579043	BI579043
98	16.8	84.0	528	12	BI365136	RE50367.5	171	16.8	84.0	599	28	A2300928	A2300928
99	16.8	84.0	528	12	BI365546	RE52109.5	172	16.8	84.0	600	12	BI168852	BI168852
100	16.8	84.0	528	12	BI365744	RE52345.5	173	16.8	84.0	600	12	BI229142	BI229142
101	16.8	84.0	528	12	BI367407	RE53168.5	174	16.8	84.0	600	12	BI364207	BI364207
102	16.8	84.0	528	12	BI374213	RE61747.5	175	16.8	84.0	600	12	BI364331	BI364331
103	16.8	84.0	529	12	BI164845	RE04596.5	176	16.8	84.0	600	12	BI578827	BI578827
104	16.8	84.0	529	12	BI230985	RE16659.5	177	16.8	84.0	601	12	BI577743	BI577743
105	16.8	84.0	529	12	BI368430	RE45196.5	178	16.8	84.0	602	9	AI259223	AI259223
106	16.8	84.0	531	12	BI215834	RE22567.5	179	16.8	84.0	602	12	BI161506	BI161506
107	16.8	84.0	531	12	BI361983	RE46430.5	180	16.8	84.0	602	12	BI235077	BI235077
108	16.8	84.0	532	12	BI370861	RE57829.5	181	16.8	84.0	602	12	BI613745	BI613745
109	16.8	84.0	534	12	BI368975	RE55044.5	182	16.8	84.0	602	12	BQ296694	BQ296694
110	16.8	84.0	535	12	BI568993	RE01228.5	183	16.8	84.0	604	9	AA440041	AA440041
111	16.8	84.0	537	12	BI484987	RE68192.5	184	16.8	84.0	604	12	BI164661	BI164661
112	16.8	84.0	538	14	CP861330	P8200071B	185	16.8	84.0	604	12	BI375713	BI375713
113	16.8	84.0	539	12	BI357238	RE43659.5	186	16.8	84.0	605	12	BI235202	BI235202
114	16.8	84.0	546	12	BI363769	RE48710.5	187	16.8	84.0	606	12	BI172172	BI172172
115	16.8	84.0	547	12	BI241414	RE38472.5	188	16.8	84.0	606	12	BI591997	BI591997
116	16.8	84.0	547	12	BI365635	RE42711.5	189	16.8	84.0	606	12	BI605603	BI605603
117	16.8	84.0	548	12	BI370055	RE56431.5	190	16.8	84.0	606	12	BI624943	BI624943
118	16.8	84.0	548	12	BI481540	RE64111.5	191	16.8	84.0	607	12	BI169813	BI169813
119	16.8	84.0	556	12	BI370600	RE57293.5	192	16.8	84.0	608	9	AA735381	AA735381
120	16.8	84.0	557	12	BI243961	RE41783.5	193	16.8	84.0	608	9	AI108250	AI108250
121	16.8	84.0	559	12	BI375475	RE63302.5	194	16.8	84.0	608	12	BI584497	BI584497
122	16.8	84.0	560	12	BI366330	RE51841.5	195	16.8	84.0	608	14	CA807327	CA807327
123	16.8	84.0	561	12	BI242099	RE39328.5	196	16.8	84.0	609	10	BE977127	BE977127
124	16.8	84.0	561	12	BI243597	RE41384.5	197	16.8	84.0	609	12	BI234434	BI234434
125	16.8	84.0	562	12	BI216600	RE23794.5	198	16.8	84.0	609	12	BI364611	BI364611
126	16.8	84.0	562	12	BI367617	RE53419.5	199	16.8	84.0	610	12	BI215547	BI215547
127	16.8	84.0	562	12	BI486118	RE69503.5	200	16.8	84.0	610	12	BI568122	BI568122
128	16.8	84.0	563	9	AA390531	LD08747.5	201	16.8	84.0	610	12	BI607657	BI607657
129	16.8	84.0	563	12	BI240070	RE36717.5	202	16.8	84.0	610	12	BI607657	BI607657
130	16.8	84.0	563	12	BI486757	RE70221.5	203	16.8	84.0	611	12	BI604848	BI604848
131	16.8	84.0	564	12	BI361738	RE46078.5	204	16.8	84.0	611	12	BI626443	BI626443
132	16.8	84.0	565	12	BI163984	RE20335.5	205	16.8	84.0	613	9	AI517455	AI517455
133	16.8	84.0	565	12	BI368351	RE254307.5	206	16.8	84.0	614	9	AI517455	AI517455
134	16.8	84.0	566	9	AA539800	LD19053.5	207	16.8	84.0	614	12	BI172007	BI172007
135	16.8	84.0	568	12	BI366714	RE52310.5	208	16.8	84.0	614	12	BI358429	BI358429
136	16.8	84.0	570	9	AI389199	GH20346.5	209	16.8	84.0	615	14	CA805518	CA805518
137	16.8	84.0	570	9	AI389199	GH20346.5	210	16.8	84.0	615	12	BI607501	BI607501
138	16.8	84.0	571	12	BI576684	RE44221.5	211	16.8	84.0	616	9	AI135561	AI135561
139	16.8	84.0	572	12	BI162252	RE01620.5	212	16.8	84.0	616	12	BI357706	BI357706
140	16.8	84.0	573	12	BI618133	RE48692.5	213	16.8	84.0	616	12	BI579025	BI579025
141	16.8	84.0	577	9	AI238765	GH14872.5	214	16.8	84.0	617	12	BI164335	BI164335
142	16.8	84.0	577	12	BI238504	RE34985.5	215	16.8	84.0	617	12	BI231890	BI231890
143	16.8	84.0	577	12	BI364356	RE49447.5	216	16.8	84.0	617	12	BI373623	BI373623
144	16.8	84.0	577	29	CG624831	OSXACS0TV	217	16.8	84.0	617	12	BI375535	BI375535
145	16.8	84.0	578	9	AI389734	GH20975.5	218	16.8	84.0	618	9	AI110215	AI110215
146	16.8	84.0	578	12	BI241945	RE39151.5	219	16.8	84.0	618	12	BI367381	BI367381
147	16.8	84.0	578	12	BI623417	RE55662.5	220	16.8	84.0	619	12	BI375379	BI375379
148	16.8	84.0	583	12	BI167321	RE07678.5	221	16.8	84.0	620	9	AI238530	AI238530
149	16.8	84.0	585	12	BI214092	RE19802.5	222	16.8	84.0	620	9	AI258403	AI258403
150	16.8	84.0	586	12	BI364504	RE49338.5	223	16.8	84.0	621	12	BI215987	BI215987

```
224 16.8 84.0 621 12 BI565840
225 16.8 84.0 623 12 BI166710
226 16.8 84.0 624 9 AI297684
227 16.8 84.0 624 12 BI364910
228 16.8 84.0 625 12 BI591996
229 16.8 84.0 626 9 AI405628
230 16.8 84.0 630 9 AI295441
231 16.8 84.0 630 10 BF498793
232 16.8 84.0 630 12 BI15419
233 16.8 84.0 630 29 CC752514
234 16.8 84.0 631 12 BI232468
235 16.8 84.0 631 12 BI580703
236 16.8 84.0 631 12 BI580810
237 16.8 84.0 632 12 BI363330
238 16.8 84.0 633 12 BI169213
239 16.8 84.0 633 12 BI172162
240 16.8 84.0 633 29 AG116253
241 16.8 84.0 634 12 BI214904
242 16.8 84.0 634 12 BI242789
243 16.8 84.0 634 12 BI629921
244 16.8 84.0 635 12 BI228617
245 16.8 84.0 635 12 BI565430
246 16.8 84.0 635 12 BI573264
247 16.8 84.0 636 12 BI169784
248 16.8 84.0 637 12 BI228554
249 16.8 84.0 637 12 BI371288
250 16.8 84.0 638 12 BI484890
251 16.8 84.0 639 9 AI238918
252 16.8 84.0 639 9 AI519051
253 16.8 84.0 639 12 BI629692
254 16.8 84.0 640 12 BI579153
255 16.8 84.0 641 12 BI365208
256 16.8 84.0 641 13 BQ232510
257 16.8 84.0 643 9 AA441085
258 16.8 84.0 644 12 BI579699
259 16.8 84.0 646 9 AI388592
260 16.8 84.0 646 12 BI237780
261 16.8 84.0 646 12 BI487106
262 16.8 84.0 646 12 BI577177
263 16.8 84.0 646 12 BI579281
264 16.8 84.0 647 9 AA736185
265 16.8 84.0 647 12 BI229262
266 16.8 84.0 647 12 BI371214
267 16.8 84.0 648 12 BI169408
268 16.8 84.0 648 12 BI216098
269 16.8 84.0 648 12 BI227540
270 16.8 84.0 648 12 BI238417
271 16.8 84.0 649 12 BI169356
272 16.8 84.0 650 9 AI063600
273 16.8 84.0 650 12 BI216602
274 16.8 84.0 650 12 BI236631
275 16.8 84.0 650 14 CA803177
276 16.8 84.0 651 12 BI213937
277 16.8 84.0 652 9 AA696210
278 16.8 84.0 652 9 AA942427
279 16.8 84.0 652 12 BI216725
280 16.8 84.0 654 12 BI375477
281 16.8 84.0 656 9 AI386508
282 16.8 84.0 656 12 BI231861
283 16.8 84.0 656 12 BI625747
284 16.8 84.0 657 12 BI234575
285 16.8 84.0 657 12 BI236602
286 16.8 84.0 657 12 BI239825
287 16.8 84.0 657 12 BI580105
288 16.8 84.0 658 12 BI228473
289 16.8 84.0 658 12 BI577865
290 16.8 84.0 658 12 BI586497
291 16.8 84.0 658 12 BI638518
292 16.8 84.0 659 12 BI167175
293 16.8 84.0 660 12 BI231162
294 16.8 84.0 661 12 BI238676
295 16.8 84.0 662 12 BI238862
296 16.8 84.0 662 12 BI372915

BI565840 RH63667.5
BI166710 RH46718.5
AI297684 LP12044.5
BI364910 RE50110.5
BI591996 RH09577.5
AI405628 GH25714.5
AI295441 LP09068.5
BF498793 AT13265.5
BI15419 RH45040.5
CC752514 ZMMBB013
BI232468 RE28539.5
BI580703 RE74926.5
BI580810 RE75047.5
BI363330 RE48110.5
BI169213 RE10323.5
BI172162 RE59049.5
AG116253 Pan trogl
BI214904 RE21054.5
BI242789 RE40217.5
BI629921 RH58860.5
BI228617 RE26214.5
BI565430 RH63205.5
BI573264 RH09957.5
BI169784 RE11056.5
BI228554 RE26126.5
BI371288 RE58152.5
BI484890 RE68064.5
AI238918 GH15057.5
AI519051 LD38175.5
BI629692 RH58594.5
BI579153 RE73139.5
BI365208 RE50450.5
BQ232510 NISC.r122
AA441085 LD15868.5
BI579699 RE73769.5
AI388592 GH19549.5
BI237780 RE34073.5
BI487106 RE70616.5
BI577177 RE70827.5
BI579281 RE73281.5
AA736185 HLO8132.5
BI229262 RE27012.5
BI371214 RE58061.5
BI169408 RE10575.5
BI216098 RE23014.5
BI227540 RE24618.5
BI238417 RE34873.5
BI169356 RE10506.5
AI063600 GH03506.5
BI216602 RE23796.5
BI236631 RE32710.5
CA803177 ESG011A.E
BI213937 RE19586.5
AA696210 GH05286.5
AA942427 LD26644.5
BI216725 RE23956.5
BI375477 RE63305.5
AI386508 GH16608.5
BI231861 RE24807.5
BI625747 RH66231.5
BI234575 RE30293.5
BI236602 RE26274.5
BI239825 RE36422.5
BI580105 RE74242.5
BI228473 RE26026.5
BI577865 RE11623.5
BI586497 RH26453.5
BI638518 SD20821.5
BI167175 RE07504.5
BI231162 RE19376.5
BI238676 RE35184.5
BI238862 RE41987.5
BI372915 RE60114.5

RESULT 1
LOCUS AG056593 1040 bp DNA linear GSS 02-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-042020.F, genomic survey sequence.
ACCESSION AG056593
VERSION AG056593.1 GI:16594052
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 1040)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..1040
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-042020.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match 85.0%; Score 17; DB 29; Length 1040;
Best Local Similarity 100.0%; Pred. No. 1.6e+04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCGCGCACCACCTCGA 17
DB 372 GCGCGCACCACCTCGA 389
RESULT 2
LOCUS BI470080 223 bp mRNA linear EST 29-NOV-2001
DEFINITION saf38ell.y3 Gm-cl077 Glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-cl077-981.5, similar to TR:Q9T0L6 Q9T0L6 HYPOTHETICAL 23.2
KD PROTEIN. ;, mRNA sequence.
ACCESSION BI470080
VERSION BI470080.1 GI:15286189
KEYWORDS EST.
SOURCE Glycine max (soybean)
```


SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 305)
 AUTHORS Simin,K., Scuderi,A., Reamey,J., Dunn,D.M., Weiss,R.B., Metherall,J.E. and Letsou,A.
 TITLE An automated screen for spatially restricted transcripts in Drosophila embryogenesis
 JOURNAL Genome Res. (2002) In press
 COMMENT Contact: Letsou, A.
 Department of Human Genetics
 Eccles Institute of Human Genetics, University of Utah
 15 N 2030 EAST RM 2100, Salt Lake City, UT 84112-5330, USA
 Tel: 801 581-4422
 Fax: 801 581-7796
 Email: aletsou@genetics.utah.edu
 Seq primer: SP6
 Location/Qualifiers
 FEATURES
 source 1..305
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /dev_stage="embryonic 8-12 hr post-fertilization"
 /clone_lib="Drosophila 8-12 hr embryonic cDNA library"
 /note="Vector: plasmid pNB; see Brown, N.H. and Kafatos, F.C. (1988) Functional cDNA libraries from Drosophila embryos. J. Mol Biol. 203, 425-37."
 ORIGIN
 Query Match 84.0%; Score 16.8; DB 13; Length 305;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCACCACTCGACTT 20
 |||||
 Db 240 GCTCCGCACCATCTCGACTT 259
 RESULT 5
 B1212140 309 bp mRNA linear EST 11-JUL-2001
 LOCUS B1212140
 DEFINITION Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 309)
 AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
 TITLE BGP/HMI RE Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BGGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 hit genomic AE003823; arm:2R [6997351..7260081]
 estimated-cyto:48D8-48F5; 04/12/2001
 Plate: RE.171 row: F column: 4
 High quality sequence stop: 276.
 Location/Qualifiers
 FEATURES
 source 1..310
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE15633"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
 ORIGIN
 Query Match 84.0%; Score 16.8; DB 12; Length 309;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCACCACTCGACTT 20
 |||||
 Db 238 GCTCCGCACCATCTCGACTT 257
 RESULT 6
 B1172928 310 bp mRNA linear EST 09-JUL-2001
 LOCUS B1172928
 DEFINITION Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 310)
 AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
 TITLE BGP/HMI RE Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BGGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 hit genomic AE003823; arm:2R [6997351..7260081]
 estimated-cyto:48D8-48F5; 04/12/2001
 Plate: RE.156 row: C column: 9
 High quality sequence stop: 276.
 Location/Qualifiers
 FEATURES
 source 1..310
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE15633"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

High quality sequence stop: 272.
 Location/Qualifiers
 FEATURES
 source 1..309
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE17164"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
 ORIGIN
 Query Match 84.0%; Score 16.8; DB 12; Length 309;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCACCACTCGACTT 20
 |||||
 Db 238 GCTCCGCACCATCTCGACTT 257
 RESULT 6
 B1172928 310 bp mRNA linear EST 09-JUL-2001
 LOCUS B1172928
 DEFINITION Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 310)
 AUTHORS Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.
 TITLE BGP/HMI RE Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: Stapleton, M.
 BGGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 hit genomic AE003823; arm:2R [6997351..7260081]
 estimated-cyto:48D8-48F5; 04/12/2001
 Plate: RE.156 row: C column: 9
 High quality sequence stop: 276.
 Location/Qualifiers
 FEATURES
 source 1..310
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE15633"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DHS-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 310;
Best Local Similarity 90.0%; Pred. No. 1.5e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20

DB 238 GCTCGCACCACTCGACTT 257

RESULT 7

BF486064 333 bp mRNA linear EST 19-APR-2001
LOCUS AT20210.5prime AT Drosophila melanogaster adult testes pOT57
DEFINITION Drosophila melanogaster cDNA clone AT20210.5 similar to gulf.
FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;
04/09/2001, mRNA sequence.

ACCESSION BF486064

VERSION BF486064.2 GI:13697956

KEYWORDS EST

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 333)

AUTHORS

Stapleton, M., Brockstein, P., Hong, L., Achavani, A., Baxter, E., Berman, B., Carlson, J., Champe, M., Chavez, C., Chev, M., Dorsett, V., Farfan, D., Frise, R., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miranda, A., Misra, S., Mungall, C.J., Nunco, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.

BDGP/HMI AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11569365.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AE003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 04/09/2001

Plate: AT.202 row: A column: 10

High quality sequence stop: 324.

Location/Qualifiers

FEATURES

source

1..333

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="AT20210"

/sex="male"

/dev_stages="0-3 day old Ore-R males"

/lab_host="Plates AT.10-AT.120: DH5-alpha. Plates

AT.121-AT.319: DH5-alpha Tona"

/clone_lib="AT Drosophila melanogaster adult testes pOT57"

/note="Organ: ADULT testes; Vector: pOT57; Site: 1: EcoRI;

Site 2: XhoI; The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

ligated into pOT57. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 333;
Best Local Similarity 90.0%; Pred. No. 1.5e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20

DB 117 GCTCGCACCACTCGACTT 136

RESULT 8

BI241833

LOCUS

DEFINITION

BI241833 333 bp mRNA linear EST 12-JUL-2001
RE39016.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE39016.5 similar to gulf.
FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;
05/12/2001, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

Stapleton, M., Brockstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, R., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunco, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.

BDGP/HMI RE Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AE003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 05/12/2001

Plate: RE.390 row: B column: 4

High quality sequence stop: 306.

Location/Qualifiers

FEATURES

source

1..333

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="RE39016"

/sex="male and female"

/dev_stages="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

/clone_lib="RE Drosophila melanogaster normalized Embryo

pFlc-1"

/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:

BamHI; Library was kindly generated by Piero Carninci at

the RIKEN. The library was normalized and excised using

Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 333;
Best Local Similarity 90.0%; Pred. No. 1.5e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20

DB 238 GCTCGCACCACTCGACTT 257

RESULT 9

BI363139

LOCUS

DEFINITION

BI363139 335 bp mRNA linear EST 01-AUG-2001
RE47880.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE47880.5 similar to gulf.
FBan0016747 located on: 2R 48F1-48F1; 05/13/2001, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

EST

```

SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 335)
AUTHORS    Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
TITLE      BDGP/HMI RE Drosophila EST Project
JOURNAL    Unpublished (2001)
COMMENT    Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003823: arm:2R [6997351..7260081]
            estimated-cyto:48D8-48F5: 05/13/2001
            Plate: RE.478 row: G column: 8
            High quality sequence stop: 334.
            Location/Qualifiers
FEATURES   source
            1..335
            /organism="Drosophila melanogaster"
            /mol_type="mRNA"
            /db_xref="taxon:7227"
            /clone="RE47880"
            /sex="male and female"
            /dev_stage="0-24 hours mixed stage embryonic"
            /lab_host="DHS-alpha Tona"
            /clone_lib="RE Drosophila melanogaster normalized Embryo
            pFlc-1"
            /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
            BamHI; Library was kindly generated by Piero Carninci at
            the RIKEN. The library was normalized and excised using
            Cre recombinase. Plasmid cDNA library."
ORIGIN
Query Match      84.0%; Score 16.8; DB 12; Length 335;
Best Local Similarity 90.0%; Pred. No. 1.5e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGCGCACCACTCGACTT 20
    |||||
DB 130 GCTCCGCACCATCTCGACTT 149
    |||||
RESULT 11
LOCUS      BI578091
DEFINITION BI578091 345 bp mRNA linear EST 06-SEP-2001
            RE71891:Spime RE Drosophila melanogaster normalized Embryo pFlc-1
            Drosophila melanogaster cDNA clone RE71891 5 similar to guf:
            FBan0016747 GO: [ornithine decarboxylase inhibitor (GO:0008073)]
            located on: 2R 48F1-48F1.; 05/18/2001, mRNA sequence.
ACCESSION  BI578091
VERSION     BI578091.1 GI:15469513
KEYWORDS    EST.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 345)
AUTHORS    Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
TITLE      BDGP/HMI RE Drosophila EST Project
JOURNAL    Unpublished (2001)
COMMENT    Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: RE.718 row: H column: 7
            High quality sequence stop: 232.
            Location/Qualifiers
FEATURES   source
            1..345
            /organism="Drosophila melanogaster"
            /mol_type="mRNA"
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 336)
AUTHORS    Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
            Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
            George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
            Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
            Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
            Rubin,G.M.
TITLE      BDGP/HMI RE Drosophila EST Project
JOURNAL    Unpublished (2001)
COMMENT    Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: RE.718 row: H column: 7
            High quality sequence stop: 232.
            Location/Qualifiers
FEATURES   source
            1..345
            /organism="Drosophila melanogaster"
            /mol_type="mRNA"

```

/db_xref="taxon:7227"
 /clone="RE71891"
 /sex="male and female"
 /dev_stages="0-24 hours mixed stage embryonic"
 /lab_host="DH5-alpha Tona"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 15.8; DB 12; Length 345;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 ||| ||||| ||||| |||||
 Db 238 GCTCCGACCATCTCGACTT 257

RESULT 12

BI227580

LOCUS

DEFINITION BI227580 346 bp mRNA linear EST 11-JUL-2001
 Drosophila melanogaster cDNA clone RE24669 5 similar to guf;
 FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1.1;
 04/12/2001, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
 Misra, S., Mungall, C.J., Nuncio, J., Pacleb, J., Paragas, V., Park, S.,
 Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
 Rubin, G.M.

TITLE BDGP/HMI RE Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AB003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 04/12/2001

Plate: RE 246 row: F column: 9

High quality sequence stop: 276.

FEATURES

source

1..346
 /organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="RE24669"

/sex="male and female"

/dev_stages="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

/clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"

/note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 346;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 ||| ||||| ||||| |||||
 Db 238 GCTCCGACCATCTCGACTT 257

RESULT 13

BI361366

LOCUS

DEFINITION BI361366 356 bp mRNA linear EST 01-AUG-2001
 Drosophila melanogaster cDNA clone RE45550 5 similar to guf;
 FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1.1;
 05/13/2001, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
 Misra, S., Mungall, C.J., Nuncio, J., Pacleb, J., Paragas, V., Park, S.,
 Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
 Rubin, G.M.

TITLE BDGP/HMI RE Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

hit genomic AB003823: arm:2R [6997351,7260081]

estimated-cyto:48D8-48F5: 05/13/2001

Plate: RE 455 row: E column: 2

High quality sequence stop: 311.

FEATURES

source

1..356
 /organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="RE45550"

/sex="male and female"

/dev_stages="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

/clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"

/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 356;
 Best Local Similarity 90.0%; Pred. No. 1.5e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 ||| ||||| ||||| |||||
 Db 130 GCTCCGACCATCTCGACTT 149

RESULT 14

AI061909

LOCUS

DEFINITION AI061909 384 bp mRNA linear EST 19-APR-2001
 LD34973.5prime LD Drosophila melanogaster embryo pOT2 Drosophila

melanogaster cDNA clone LD34973 5prime similar to AF038597;
Drosophila melanogaster ornithine decarboxylase antizyme mRNA,
complete cds, mRNA sequence.

ACCESSION

AI061909

VERSION

AI061909.1 GI:3337748

KEYWORDS

EST

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 364)

AUTHORS

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Teang, G.,
Lewis, S. and Rubin, G.M.

TITLE

BGCP/HHMI Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

Contact: Stapleton, M.

FEATURES

Location/Qualifiers

1..364

source

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="LD34973"

/sex="male and female"

/dev_stage="0 to 24 hours mixed stage embryonic"

/lab_host="X11 Blue"

/clone_lib="LD Drosophila melanogaster embryo pOR2"

/note="Organ: embryo; Vector: pOR2; Site:1; EcoRI; Site:2;
XhoI; Sized fractionated cDNAs were directly ligated into
pOR2."

ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 364;

Best Local Similarity 90.0%; Pred. No. 1.5e+04;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy

1 GCGCGCACCACCTCGACTT 20

Db

63 GCTCCGCACCATCTCGACTT 82

RESULT 15

AA201741

LOCUS

AA201741

DEFINITION

LD04952.5prime LD Drosophila melanogaster embryo BlueScript

ACCESSION

AA201741

VERSION

AA201741.1 GI:2777292

KEYWORDS

EST

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 378)

AUTHORS

Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Teang, G.,
Lewis, S. and Rubin, G.M.

TITLE

BGCP/HHMI Drosophila EST Project

JOURNAL

Unpublished (2001)

COMMENT

On Jan 14, 1998 this sequence version replaced gi:1797526.

CONTACT

Contact: Stapleton, M.

BGP

Lawrence Berkeley National Lab

One Cyclotron Rd,

Berkeley, CA 94720, USA

Fax:

510 486 6798

Email:

http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

Blast expect value = 5.0E-87 on U29529: Drosophila melanogaster
gutfeeling (guf) mRNA, complete cds
Plate: 49 row: E column: 4
High quality sequence stop: 305.

FEATURES

Location/Qualifiers

1..378

source

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="BDGP_EST:BDcln004335"

/db_xref="taxon:7227"

/clone="LD04952"

/sex="male and female"

/dev_stage="0 to 24 hours mixed stage embryonic"

/lab_host="SOLR"

/clone_lib="LD Drosophila melanogaster embryo Bluescript"

/note="Organ: embryo; Vector: Bluescript SK; Site:1;
EcoRI; Site:2; XhoI; Constructed using Stratagene ZAP-cDNA
Synthesis kit. Oligo dt-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"

ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 378;

Best Local Similarity 90.0%; Pred. No. 1.5e+04;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy

1 GCGCGCACCACCTCGACTT 20

Db

201 GTCGCGCACCACCTCGACTT 220

RESULT 16

BI243125

LOCUS

BI243125

DEFINITION

393 bp mRNA linear EST 12-JUL-2001

RE40682.5prime RE Drosophila melanogaster normalized Embryo pFLC-1

Drosophila melanogaster cDNA clone RE40682 5 similar to guf;

FBanc016747; enzyme inhibitor; located on: 2R 48F1-48F1;

05/12/2001, mRNA sequence.

BI243125

BI243125.1 GI:14711979

EST

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 393)

Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,

Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,

George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,

Phuanavanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and

Rubin, G.M.

BDGP/HHMI RE Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu

hit genomic AE003823; arm:28 [697351,7260081]

estimated-cyto:48D8-48F5: 05/12/2001

Plate: RE.406 row: G column: 10

High quality sequence stop: 322.

Location/Qualifiers

1..393

/organism="Drosophila melanogaster"

/mol_type="mRNA"

/db_xref="taxon:7227"

/clone="RE40682"

/sex="male and female"

/dev_stage="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

KEYWORDS EST
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

REFERENCE
1 (bases 1 to 399)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.

TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.

BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003823: arm:2R [6997351,7260081]
estimated-cyto:48D8-48F5: 05/13/2001
Plate: RE.421 row: B column: 2
High quality sequence stop: 346.

FEATURES
source
1. 399
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="R842114"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
/note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 84.0%; Score 16.8; DB 12; Length 399;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCGCGCGCACCATCTCGACTT 20
|||
Db 238 GCTCCGACCATCTCGACTT 257

RESULT 20
BI357752 404 bp mRNA linear EST 31-JUL-2001
LOCUS R844306.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
DEFINITION Drosophila melanogaster cDNA clone R844306 5 similar to guf:
F8an0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;;
05/13/2001, mRNA sequence.

ACCESSION BI357752
VERSION BI357752.1 GI:15052206
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

REFERENCE
1 (bases 1 to 404)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.

TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.

BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003823: arm:2R [6997351,7260081]
estimated-cyto:48D8-48F5: 05/13/2001
Plate: RE.443 row: A column: 6
High quality sequence stop: 292.

FEATURES
source
1. 404
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE44306"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"
/note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 84.0%; Score 16.8; DB 12; Length 404;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCGCGCGCACCATCTCGACTT 20
|||
Db 237 GCTCCGACCATCTCGACTT 256

RESULT 21
AA695426 414 bp mRNA linear EST 23-APR-2001
LOCUS GM02813.5prime GM Drosophila melanogaster ovary Bluescript
DEFINITION Drosophila melanogaster cDNA clone GM02813 5prime, mRNA sequence.

ACCESSION AA695426
VERSION AA695426.1 GI:2698046
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

REFERENCE
1 (bases 1 to 414)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

AUTHORS
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S. and Rubin, G.M.

TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.

BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 28 row: B column: 1
High quality sequence stop: 206.

FEATURES
source
1. 414
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="GM02813"
/sex="female"
/dev_stage="newly enclosed females: germarium-stage 6"
/lab_host="SOLR"

/clone lib="GM Drosophila melanogaster ovary Bluescript"
 /note="organ: ovary; Vector: Bluescript SK; Site 1: EcoRI;
 Site 2: XhoI; Constructed using Stratagene ZAP-cDNA
 Synthesis kit. Oligo dt-primed and directionally cloned at
 EcoRI and XhoI in Bluescript SK(+/-)"

ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 414;
 Best Local Similarity 90.0%; Pred. No. 1.6e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 DB 212 GCTCCGCACCACTCGACTT 231

RESULT 22

B1577942 414 bp mRNA linear EST 06-SEP-2001
 LOCUS RE17118.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
 DEFINITION Drosophila melanogaster cDNA clone RE17118 5 similar to gtf:
 FBan0016747 GO: [ornithine decarboxylase inhibitor (GO:0008073)]
 located on: 2R 48F1-48F1; 05/18/2001, mRNA sequence.

ACCESSION B1577942.1 GI:15469364
 VERSION
 KEYWORDS
 SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 414)
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
 George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
 Misra, S., Mungall, C. J., Nuno, J., Pacle, J., Paragas, V., Park, S.,
 Phuanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and
 Rubin, G. M.

TITLE BDGP/HMI RE Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 hit genomic A5003823: arm:2R [6997351,7260081]
 estimated-cyto:48D8-48F5: 05/18/2001
 Plate: RE.717 row: B column: 6
 High quality sequence stop: 308.

FEATURES

source

1..414
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /clone="RE17118"
 /sex="male and female"
 /dev_stage="0-24 hours mixed stage embryonic"
 /lab_host="DH5-alpha TonA"
 /clone_lib="RE Drosophila melanogaster normalized Embryo
 pFlc-1"
 /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
 BamHI; Library was kindly generated by Piero Carninci at
 the RIKEN. The library was normalized and excised using
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 414;
 Best Local Similarity 90.0%; Pred. No. 1.6e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||

Db 238 GCTCCGCACCACTCGACTT 257

RESULT 23
 BQ103130

LOCUS BQ103130 418 bp mRNA linear EST 12-APR-2002
 DEFINITION UGC0528 Drosophila 8-12 hr embryonic cDNA library Drosophila
 melanogaster cDNA 5', mRNA sequence.

ACCESSION BQ103130
 VERSION BQ103130.1 GI:20144613
 KEYWORDS EST

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 418)
 AUTHORS Simin, K., Scuderi, A., Reamey, J., Dunn, D. M., Weiss, R. B.,
 Wetherall, J. E. and Letsou, A.
 TITLE An automated screen for spatially restricted transcripts in
 Drosophila embryogenesis

JOURNAL Genome Res. (2002) In press

COMMENT Contact: Letsou, A.

Department of Human Genetics

Eccles Institute of Human Genetics, University of Utah

15 N 2030 EAST RM 2100, Salt Lake City, UT 84112-5330, USA

Tel: 801 581-4422

Fax: 801 581-7796

Email: aletsou@genetics.utah.edu

Seq primer: SP6.

FEATURES

source

1..418
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /dev_stage="embryonic 8-12 hr post-fertilization"
 /clone_lib="Drosophila 8-12 hr embryonic cDNA library"
 /note="Vector: plasmid pNB; see Brown, N.H. and Kafatos,
 F.C. (1988) Functional cDNA libraries from Drosophila
 embryos. J. Mol Biol. 203, 425-37."

ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 418;
 Best Local Similarity 90.0%; Pred. No. 1.6e+04;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20

|||||

Db 194 GCTCCGCACCACTCGACTT 213

RESULT 24

BQ103552 419 bp mRNA linear EST 12-APR-2002
 LOCUS UGC0959 Drosophila 8-12 hr embryonic cDNA library Drosophila
 melanogaster cDNA 5', mRNA sequence.

ACCESSION BQ103552

VERSION BQ103552.1 GI:20145035

KEYWORDS EST

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 419)

AUTHORS Simin, K., Scuderi, A., Reamey, J., Dunn, D. M., Weiss, R. B.,
 Wetherall, J. E. and Letsou, A.

TITLE An automated screen for spatially restricted transcripts in

Drosophila embryogenesis

JOURNAL Genome Res. (2002) In press

COMMENT Contact: Letsou, A.

Department of Human Genetics

Eccles Institute of Human Genetics, University of Utah

pFIC-1" /note="Organ: embryo; Vector: pFIC1; Site_1: xhoI; Site_2: BamHI; Library was kindly generated by Piero Carninci at

```

RESULT 27
BI375669
LOCUS
DEFINITION
  RE63532.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE63532 5 similar to guf:
  FBan0016747 GO: [ornithine decarboxylase inhibitor (GO:0008073)]
  located on: 2R 48F1-48F1; 05/16/2001, mRNA sequence.
ACCESSION
BI375669
VERSION
BI375669.1 GI:15071697
KEYWORDS
EST.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 427)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AB003823: arm:2R [6997351,7260081]
  estimated-cyto:48D8-48F5; 05/16/2001
  Plate: RE 635 row: C column: 8
  High quality sequence stop: 426.
  Location/Qualifiers
  1..427
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /clone="RE63532"
  /sex="male and female"
  /dev_stage="0-24 hours mixed stage embryonic"
  /lab_host="DH5-alpha Tona"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pFlc-1"
  /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
  BamHI; Library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 84.0%; Score 16.8; DB 12; Length 427;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20
DB 238 GCTCCGACCACTCGACTT 257

RESULT 28
BI484665
LOCUS
DEFINITION
  RE67806.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE67806 5 similar to guf:
  FBan0016747 GO: [ornithine decarboxylase inhibitor (GO:0008073)]
  located on: 2R 48F1-48F1; 05/22/2001, mRNA sequence.
ACCESSION
BI484665
VERSION
BI484665.1 GI:15324957
KEYWORDS
EST.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 427)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AB003823: arm:2R [6997351,7260081]
  estimated-cyto:48D8-48F5; 05/16/2001
  Plate: RE 635 row: C column: 8
  High quality sequence stop: 426.
  Location/Qualifiers
  1..427
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /clone="RE63532"
  /sex="male and female"
  /dev_stage="0-24 hours mixed stage embryonic"
  /lab_host="DH5-alpha Tona"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pFlc-1"
  /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
  BamHI; Library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 84.0%; Score 16.8; DB 12; Length 427;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20
DB 238 GCTCCGACCACTCGACTT 257

```

```

REFERENCE
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 427)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AB003823: arm:2R [6997351,7260081]
  estimated-cyto:48D8-48F5; 05/22/2001
  Plate: RE 678 row: A column: 6
  High quality sequence stop: 379.
  Location/Qualifiers
  1..427
  /organism="Drosophila melanogaster"
  /mol_type="mRNA"
  /db_xref="taxon:7227"
  /clone="RE67806"
  /sex="male and female"
  /dev_stage="0-24 hours mixed stage embryonic"
  /lab_host="DH5-alpha Tona"
  /clone_lib="RE Drosophila melanogaster normalized Embryo
  pFlc-1"
  /note="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
  BamHI; Library was kindly generated by Piero Carninci at
  the RIKEN. The library was normalized and excised using
  Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 84.0%; Score 16.8; DB 12; Length 427;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCCGACCACTCGACTT 20
DB 237 GCTCCGACCACTCGACTT 256

RESULT 29
BI365791
LOCUS
DEFINITION
  RE51171.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE51171 5 similar to guf:
  FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;
  05/14/2001, mRNA sequence.
ACCESSION
BI365791
VERSION
BI365791.1 GI:15061819
KEYWORDS
EST.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 428)
  Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
  Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
  George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
  Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
  Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and
  Rubin,G.M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP

```

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AB003823: arm:2R [6997351,7260081]
estimated-cyto:48D8-48F5: 05/14/2001
Plate: RE.511 row: F column: 11
High quality sequence stop: 426.
Location/Qualifiers
1. .428
/organism="Drosophila melanogaster"
/mol_type="mRNA:7227"
/db_xref="taxon:7227"
/clone="RE51171"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/notes="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site: 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 428;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGACCACTCGACTT 20
|||
Db 238 GCTCGGACCACTCGACTT 257

RESULT 30
BI366732 428 bp mRNA linear EST 01-AUG-2001
LOCUS
DEFINITION
RE52331.1 Sprime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster cDNA clone RE52331 5 similar to gcf:
FBan0016747 'enzyme inhibitor' located on: 2R 48F1-48F1;
05/14/2001, mRNA sequence.

ACCESSION
BI366732
VERSION
BI366732.1 GI:15062760
KEYWORDS
EST.
SOURCE
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 428)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
Rubin, G.M.
BDGP/HMI RE Drosophila EST Project
Unpublished (2001)
CONTACT: Stapleton, M.
BDGP

TITLE
JOURNAL
COMMENT
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE003823: arm:2R [6997351,7260081]
estimated-cyto:48D8-48F5: 05/14/2001
Plate: RE.523 row: C column: 7
High quality sequence stop: 426.
Location/Qualifiers
1. .428
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"

FEATURES
source

/clone="RE52331"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/note="Organ: embryo; Vector: pFlc1; Site: 1: XhoI; Site: 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 428;
Best Local Similarity 90.0%; Pred. No. 1.6e+04;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGACCACTCGACTT 20
|||
Db 238 GCTCGGACCACTCGACTT 257

Search completed: June 20, 2004, 14:13:52
Job time : 951.603 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 22.1164 Seconds

(without alignments)

501.846 Million cell updates/sec

Title: US-10-624-714-13

Perfect score: 20

Sequence: 1 ggcgcgcacacacacgactt 20

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

Issued Patents NA:*

1: /cgm2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgm2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgm2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgm2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgm2_6/ptodata/2/ina/PTGUS_COMB.seq.*
6: /cgm2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	20	100.0	4403765	3	US-09-103-840A-2
2	20	100.0	4411529	3	US-09-103-840A-1
3	16.4	82.0	3600	4	US-09-657-042A-3
4	15.8	79.0	987	4	US-09-252-991A-7285
5	15.8	79.0	1083	4	US-09-489-039A-2785
6	15.8	79.0	1368	4	US-09-252-991A-7174
7	15.8	79.0	1575	4	US-09-252-991A-7330
8	15.4	77.0	114	4	US-09-536-977-37
9	15.4	77.0	272	4	US-09-536-977-59
10	15.4	77.0	453	4	US-09-252-991A-12478
11	15.4	77.0	495	4	US-09-252-991A-12406
12	15.4	77.0	525	4	US-09-252-991A-12111
13	15.4	77.0	690	4	US-09-252-991A-12360
14	15.4	77.0	924	4	US-09-252-991A-12250
15	15.4	77.0	1011	4	US-09-252-991A-12311
16	15.4	77.0	1038	4	US-09-536-977-75
17	15.4	77.0	1085	4	US-09-522-714-15
18	15.4	77.0	1143	4	US-09-252-991A-12185
19	15.4	77.0	1404	4	US-09-252-991A-13158
20	15.4	77.0	2163	4	US-09-252-991A-15435
21	15.4	77.0	2469	4	US-09-536-977-71
22	15.4	77.0	2844	4	US-09-252-991A-12788
23	15.4	77.0	3342	4	US-09-252-991A-12644
24	15.4	77.0	6972	4	US-09-595-684B-38
25	15.4	77.0	7011	4	US-09-268-153-9
26	15.4	77.0	8309	4	US-09-620-312D-1083
27	15.4	77.0	4403765	3	US-09-103-840A-2

C 28	15.4	77.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 29	15.2	76.0	333	4	US-09-621-976-1316	Sequence 1116, Ap
C 30	15.2	76.0	588	4	US-09-252-991A-8968	Sequence 8968, Ap
C 31	15.2	76.0	606	4	US-09-252-991A-9120	Sequence 9120, Ap
C 32	15.2	76.0	627	4	US-09-252-991A-9038	Sequence 9038, Ap
C 33	15.2	76.0	673	4	US-09-833-381-800	Sequence 800, App
C 34	15.2	76.0	945	4	US-09-489-039A-2349	Sequence 2349, Ap
C 35	15.2	76.0	1014	4	US-09-489-039A-6032	Sequence 6032, Ap
C 36	15.2	76.0	1029	4	US-09-252-991A-153	Sequence 153, App
C 37	15.2	76.0	1038	4	US-09-252-991A-13643	Sequence 13643, A
C 38	15.2	76.0	1110	4	US-09-252-991A-173	Sequence 173, App
C 39	15.2	76.0	1161	4	US-09-252-991A-9202	Sequence 9202, Ap
C 40	15.2	76.0	1209	4	US-09-252-991A-8857	Sequence 8857, Ap
C 41	15.2	76.0	1221	6	5212396-16	Patent No. 5212396
C 42	15.2	76.0	1251	4	US-09-252-991A-13795	Sequence 13795, A
C 43	15.2	76.0	1266	4	US-09-252-991A-183	Sequence 183, App
C 44	15.2	76.0	1611	4	US-09-252-991A-183	Patent No. 5212396
C 45	15.2	76.0	1879	6	5212396-5	Sequence 5, Appli
C 46	15.2	76.0	2932	3	US-08-999-774A-5	Sequence 165, App
C 47	15.2	76.0	2933	4	US-09-149-476-185	Sequence 296, App
C 48	15.2	76.0	3276	4	US-09-149-476-298	Sequence 298, App
C 49	15.2	76.0	3768	4	US-09-566-921-50	Sequence 50, Appl
C 50	15.2	76.0	7812	3	US-09-368-590-1	Sequence 1, Appli
C 51	15.2	76.0	536165	4	US-09-214-808-1	Sequence 1, Appli
C 52	14.8	74.0	30	3	US-08-817-787-30	Sequence 30, Appl
C 53	14.8	74.0	41	4	US-09-463-067-19	Sequence 19, Appl
C 54	14.8	74.0	182	3	US-08-818-112-29	Sequence 29, Appl
C 55	14.8	74.0	182	4	US-08-818-111-29	Sequence 29, Appl
C 56	14.8	74.0	182	4	US-09-056-556-29	Sequence 29, Appl
C 57	14.8	74.0	182	4	US-09-072-596-29	Sequence 29, Appl
C 58	14.8	74.0	182	4	US-09-072-967-29	Sequence 29, Appl
C 59	14.8	74.0	272	4	US-09-313-294A-2605	Sequence 2605, Ap
C 60	14.8	74.0	459	4	US-09-489-039A-3617	Sequence 3617, Ap
C 61	14.8	74.0	663	4	US-09-252-991A-1065	Sequence 1065, Ap
C 62	14.8	74.0	687	4	US-09-252-991A-11810	Sequence 11810, A
C 63	14.8	74.0	814	4	US-09-187-999-18	Sequence 18, Appl
C 64	14.8	74.0	870	4	US-09-252-991A-15989	Sequence 15989, A
C 65	14.8	74.0	1062	4	US-09-252-991A-16198	Sequence 16198, A
C 66	14.8	74.0	1212	4	US-09-489-039A-4706	Sequence 4706, Ap
C 67	14.8	74.0	1233	4	US-09-489-039A-4691	Sequence 4691, Ap
C 68	14.8	74.0	1301	2	US-08-641-314C-1	Sequence 1, Appli
C 69	14.8	74.0	1416	4	US-09-252-991A-1100	Sequence 1100, Ap
C 70	14.8	74.0	1425	4	US-09-252-991A-11513	Sequence 11513, A
C 71	14.8	74.0	1434	4	US-09-489-039A-3399	Sequence 3399, Ap
C 72	14.8	74.0	1458	4	US-09-489-039A-1373	Sequence 1373, Ap
C 73	14.8	74.0	1470	4	US-09-252-991A-896	Sequence 896, App
C 74	14.8	74.0	1511	2	US-08-809-763-3	Sequence 3, Appli
C 75	14.8	74.0	1511	3	US-08-956-253-3	Sequence 3, Appli
C 76	14.8	74.0	1512	4	US-09-252-991A-1138	Sequence 1138, Ap
C 77	14.8	74.0	1555	2	US-08-696-376-1	Sequence 1, Appli
C 78	14.8	74.0	4326	4	US-09-252-991A-11572	Sequence 11572, A
C 79	14.8	74.0	4473	4	US-09-252-991A-11871	Sequence 11871, A
C 80	14.8	74.0	33529	3	US-09-144-085-3	Sequence 3, Appli
C 81	14.8	74.0	36033	4	US-08-311-731A-124	Sequence 124, App
C 82	14.8	74.0	68750	3	US-09-335-409-1	Sequence 1, Appli
C 83	14.8	74.0	68750	4	US-09-568-102-1	Sequence 1, Appli
C 84	14.8	74.0	68750	4	US-09-567-969-1	Sequence 1, Appli
C 85	14.8	74.0	68750	4	US-09-568-480-1	Sequence 1, Appli
C 86	14.8	74.0	68750	4	US-09-568-486-1	Sequence 1, Appli
C 87	14.8	74.0	68750	4	US-09-568-472-1	Sequence 1, Appli
C 88	14.8	74.0	68750	4	US-09-567-899-1	Sequence 1, Appli
C 89	14.8	74.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
C 90	14.4	72.0	45	2	US-08-406-855A-9	Sequence 9, Appli
C 91	14.4	72.0	45	3	US-09-206-899-9	Sequence 9, Appli
C 92	14.4	72.0	45	4	US-09-688-415-14	Sequence 14, Appl
C 93	14.4	72.0	376	4	US-09-621-976-16394	Sequence 16394, A
C 94	14.4	72.0	399	4	US-09-489-039A-6019	Sequence 6019, Ap
C 95	14.4	72.0	406	4	US-09-621-976-16393	Sequence 16393, A
C 96	14.4	72.0	477	4	US-09-252-991A-10565	Sequence 10565, A
C 97	14.4	72.0	478	4	US-09-280-116-174	Sequence 174, App
C 98	14.4	72.0	483	4	US-09-621-976-2791	Sequence 2791, Ap
C 99	14.4	72.0	516	4	US-09-252-991A-16362	Sequence 16362, A
C 100	14.4	72.0	810	4	US-09-489-039A-1365	Sequence 1365, Ap

C 101	14.4	72.0	843	4	US-08-426-630-3	Sequence 3, Appli	174	14.2	71.0	490	2	US-08-967-101-76	Sequence 76, Appl
C 102	14.4	72.0	846	4	US-09-252-991A-16471	Sequence 16471, A	175	14.2	71.0	490	2	US-08-552-541-76	Sequence 76, Appl
C 103	14.4	72.0	885	4	US-09-252-991A-16130	Sequence 16130, A	176	14.2	71.0	490	3	US-09-124-698-76	Sequence 76, Appl
C 104	14.4	72.0	906	4	US-09-252-991A-16170	Sequence 16170, A	177	14.2	71.0	490	3	US-09-127-480-76	Sequence 76, Appl
C 105	14.4	72.0	912	4	US-09-252-991A-14700	Sequence 14700, A	178	14.2	71.0	490	3	US-08-496-841C-76	Sequence 76, Appl
C 106	14.4	72.0	915	4	US-09-252-991A-14700	Sequence 14700, A	179	14.2	71.0	490	4	US-09-124-523-76	Sequence 76, Appl
C 107	14.4	72.0	1005	1	US-08-483-385A-2	Sequence 2, Appli	180	14.2	71.0	490	4	US-08-636-796A-76	Sequence 76, Appl
C 108	14.4	72.0	1035	3	US-08-733-837B-1	Sequence 1, Appli	181	14.2	71.0	490	4	US-08-431-048F-76	Sequence 76, Appl
C 109	14.4	72.0	1049	4	US-09-280-116-176	Sequence 176, App	182	14.2	71.0	738	4	US-09-134-000C-142	Sequence 142, App
C 110	14.4	72.0	1056	4	US-09-252-991A-1419	Sequence 1419, App	183	14.2	71.0	762	4	US-09-489-039A-6180	Sequence 6180, App
C 111	14.4	72.0	1125	4	US-09-252-991A-14568	Sequence 14568, A	184	14.2	71.0	883	4	US-09-561-756-20	Sequence 20, Appl
C 112	14.4	72.0	1146	4	US-09-257-580-1	Sequence 1, Appli	185	14.2	71.0	883	4	US-09-227-721-20	Sequence 20, Appl
C 113	14.4	72.0	1146	4	US-09-257-580-4	Sequence 4, Appli	186	14.2	71.0	883	4	US-09-954-697-20	Sequence 20, Appl
C 114	14.4	72.0	1239	4	US-09-489-039A-5762	Sequence 5762, App	187	14.2	71.0	894	3	US-08-651-136C-3	Sequence 3, Appli
C 115	14.4	72.0	1248	4	US-09-253-991A-15057	Sequence 15057, A	188	14.2	71.0	894	3	US-09-229-911A-3	Sequence 3, Appli
C 116	14.4	72.0	1400	4	US-09-023-655-535	Sequence 535, App	189	14.2	71.0	927	3	US-08-651-136C-5	Sequence 5, Appli
C 117	14.4	72.0	1521	4	US-09-483-039A-5902	Sequence 5902, App	190	14.2	71.0	927	4	US-09-229-911A-5	Sequence 5, Appli
C 118	14.4	72.0	1584	4	US-09-252-991A-14522	Sequence 14522, A	191	14.2	71.0	960	3	US-08-651-136C-1	Sequence 1, Appli
C 119	14.4	72.0	1593	4	US-09-252-991A-1519	Sequence 1519, App	192	14.2	71.0	960	4	US-09-229-911A-1	Sequence 1, Appli
C 120	14.4	72.0	1619	4	US-09-388-413B-5	Sequence 5, Appli	193	14.2	71.0	996	4	US-09-252-991A-5415	Sequence 5415, App
C 121	14.4	72.0	1621	1	US-08-722-001-13	Sequence 13, Appli	194	14.2	71.0	1059	4	US-09-489-039A-3031	Sequence 3031, App
C 122	14.4	72.0	1632	1	US-07-959-941-1	Sequence 1, Appli	195	14.2	71.0	1119	4	US-09-252-991A-15042	Sequence 15042, A
C 123	14.4	72.0	1632	1	US-08-259-924-1	Sequence 1, Appli	196	14.2	71.0	1143	4	US-09-480-297A-11	Sequence 11, Appli
C 124	14.4	72.0	1653	4	US-09-345-469-2	Sequence 2, Appli	197	14.2	71.0	1258	4	US-09-724-864-2	Sequence 2, Appli
C 125	14.4	72.0	1660	4	US-09-461-325-129	Sequence 129, App	198	14.2	71.0	1458	4	US-09-252-991A-8725	Sequence 8725, App
C 126	14.4	72.0	1660	4	US-10-012-542-129	Sequence 129, App	199	14.2	71.0	1572	4	US-09-253-991A-14586	Sequence 14586, A
C 127	14.4	72.0	1716	4	US-09-620-312D-859	Sequence 859, App	200	14.2	71.0	1580	1	US-08-234-783-3	Sequence 3, Appli
C 128	14.4	72.0	1723	4	US-09-461-325-98	Sequence 98, App	201	14.2	71.0	1680	1	US-08-456-907-3	Sequence 3, Appli
C 129	14.4	72.0	1723	4	US-10-011-542-98	Sequence 98, App	202	14.2	71.0	1809	5	PCT-US95-05523-3	Sequence 3, Appli
C 130	14.4	72.0	1776	1	US-08-722-001-29	Sequence 29, Appli	203	14.2	71.0	1809	5	PCT-US95-10973A-79	Sequence 79, Appli
C 131	14.4	72.0	1806	4	US-09-252-991A-14654	Sequence 14654, A	204	14.2	71.0	1905	1	US-07-832-855-3	Sequence 3, Appli
C 132	14.4	72.0	1965	4	US-09-252-991A-15096	Sequence 15096, A	205	14.2	71.0	1929	4	US-09-252-991A-8827	Sequence 8827, App
C 133	14.4	72.0	2002	4	US-08-016-434-1172	Sequence 1172, App	206	14.2	71.0	1950	3	US-08-685-466C-1	Sequence 1, Appli
C 134	14.4	72.0	2140	1	US-09-334-698-1	Sequence 1, Appli	207	14.2	71.0	2079	4	US-09-253-991A-9222	Sequence 9222, App
C 135	14.4	72.0	2140	1	US-08-228-932-1	Sequence 1, Appli	208	14.2	71.0	2400	3	US-08-930-001-1	Sequence 1, Appli
C 136	14.4	72.0	2140	1	US-08-468-939-1	Sequence 1, Appli	209	14.2	71.0	2457	4	US-09-252-991A-15181	Sequence 15181, App
C 137	14.4	72.0	2140	2	US-08-408-855A-1	Sequence 1, Appli	210	14.2	71.0	2822	4	US-09-620-312D-702	Sequence 702, App
C 138	14.4	72.0	2140	2	US-08-224-130-1	Sequence 1, Appli	211	14.2	71.0	2830	1	US-07-882-292-1	Sequence 1, Appli
C 139	14.4	72.0	2140	3	US-08-244-354-1	Sequence 1, Appli	212	14.2	71.0	2830	1	US-08-331-644-1	Sequence 1, Appli
C 140	14.4	72.0	2140	3	US-09-206-899-1	Sequence 1, Appli	213	14.2	71.0	2830	5	PCT-US93-04102-1	Sequence 2, Appli
C 141	14.4	72.0	2140	4	US-09-444-783-1	Sequence 1, Appli	214	14.2	71.0	2921	3	US-08-904-032-1	Sequence 2, Appli
C 142	14.4	72.0	2140	4	US-09-688-415-1	Sequence 1, Appli	215	14.2	71.0	3020	4	US-09-220-132-19	Sequence 19, Appl
C 143	14.4	72.0	2140	4	US-09-444-783-1	Sequence 1, Appli	216	14.2	71.0	3141	2	US-08-658-665-66	Sequence 66, Appl
C 144	14.4	72.0	2140	4	PCT-US95-04203-1	Sequence 1, Appli	217	14.2	71.0	3141	3	US-08-796-101-30	Sequence 30, Appl
C 145	14.4	72.0	2157	4	US-09-252-991A-10757	Sequence 10757, A	218	14.2	71.0	3141	3	US-09-085-273-66	Sequence 66, Appl
C 146	14.4	72.0	2253	4	US-09-252-991A-10404	Sequence 10404, A	219	14.2	71.0	3141	3	US-09-916-963-66	Sequence 66, Appl
C 147	14.4	72.0	2406	4	US-09-252-991A-1657	Sequence 1657, App	220	14.2	71.0	3288	2	US-08-620-694A-1	Sequence 1, Appli
C 148	14.4	72.0	2728	1	US-08-482-385A-5	Sequence 5, Appli	221	14.2	71.0	3288	3	US-09-022-255-1	Sequence 1, Appli
C 149	14.4	72.0	2982	1	US-08-348-143-2	Sequence 2, Appli	222	14.2	71.0	3288	3	US-09-022-255-1	Sequence 1, Appli
C 150	14.4	72.0	2982	1	US-08-571-785-2	Sequence 2, Appli	223	14.2	71.0	3288	3	US-09-022-253-1	Sequence 1, Appli
C 151	14.4	72.0	2982	4	US-09-182-435-2	Sequence 2, Appli	224	14.2	71.0	3288	3	US-09-022-253-1	Sequence 1, Appli
C 152	14.4	72.0	2982	4	US-08-558-340-2	Sequence 2, Appli	225	14.2	71.0	3288	3	US-09-022-253-1	Sequence 1, Appli
C 153	14.4	72.0	3627	4	US-08-348-143-3	Sequence 3, Appli	226	14.2	71.0	3288	3	US-09-022-253-1	Sequence 1, Appli
C 154	14.4	72.0	4027	1	US-08-348-143-3	Sequence 3, Appli	227	14.2	71.0	3288	3	US-09-022-257-1	Sequence 1, Appli
C 155	14.4	72.0	4027	1	US-08-571-785-3	Sequence 3, Appli	228	14.2	71.0	3288	4	US-09-549-679-1	Sequence 1, Appli
C 156	14.4	72.0	4027	1	US-08-571-785-3	Sequence 3, Appli	229	14.2	71.0	3567	2	US-08-658-665-69	Sequence 69, Appl
C 157	14.4	72.0	4027	1	US-08-571-785-3	Sequence 3, Appli	230	14.2	71.0	3567	3	US-08-798-101-33	Sequence 33, Appl
C 158	14.4	72.0	4027	4	US-09-182-435-3	Sequence 3, Appli	231	14.2	71.0	3567	3	US-09-085-273-69	Sequence 69, Appl
C 159	14.4	72.0	4027	4	US-09-558-340-3	Sequence 3, Appli	232	14.2	71.0	3567	4	US-09-916-963-69	Sequence 69, Appl
C 160	14.4	72.0	4027	4	US-09-558-340-3	Sequence 3, Appli	233	14.2	71.0	3994	4	US-09-738-946-7	Sequence 7, Appli
C 161	14.4	72.0	4027	4	US-09-558-340-4	Sequence 4, Appli	234	14.2	71.0	4075	2	US-08-658-665-67	Sequence 67, Appl
C 162	14.4	72.0	5398	4	US-08-426-630-1	Sequence 1, Appli	235	14.2	71.0	4075	3	US-08-796-101-31	Sequence 31, Appl
C 163	14.4	72.0	18994	1	US-08-459-586-4	Sequence 4, Appli	236	14.2	71.0	4075	3	US-09-085-273-67	Sequence 67, Appl
C 164	14.4	72.0	18994	2	US-08-282-696-4	Sequence 4, Appli	237	14.2	71.0	4075	4	US-09-915-963-67	Sequence 67, Appl
C 165	14.2	71.0	41	5	PCT-US95-10973A-55	Sequence 55, App	238	14.2	71.0	4380	4	US-08-955-585A-3	Sequence 3, Appli
C 166	14.2	71.0	56	5	PCT-US95-10973A-56	Sequence 56, App	239	14.2	71.0	4909	2	US-08-658-665-68	Sequence 68, Appl
C 167	14.2	71.0	285	4	US-08-743-294A-6726	Sequence 6726, App	240	14.2	71.0	4909	3	US-08-796-101-32	Sequence 32, Appl
C 168	14.2	71.0	330	2	US-08-743-294A-6726	Sequence 6726, App	241	14.2	71.0	4909	4	US-09-085-273-68	Sequence 68, Appl
C 169	14.2	71.0	417	4	US-09-283-144-1	Sequence 1, Appli	242	14.2	71.0	4909	3	US-09-916-963-68	Sequence 68, Appl
C 170	14.2	71.0	453	3	US-09-060-756-109	Sequence 109, App	243	14.2	71.0	5046	4	US-09-548-938A-5	Sequence 5, Appli
C 171	14.2	71.0	453	4	US-09-670-314-109	Sequence 109, App	244	14.2	71.0	6360	4	US-09-171-659-9	Sequence 9, Appli
C 172	14.2	71.0	457	3	US-08-952-664-10	Sequence 10, Appli	245	14.2	71.0	6749	2	US-08-658-665-71	Sequence 71, Appl
C 173	14.2	71.0	457	3	US-09-487-874-10	Sequence 10, Appli	246	14.2	71.0	6749	2	US-08-658-665-177	Sequence 177, App

247 14.2 71.0 6749 3 US-08-796-101-35 Sequence 35, Appl
248 14.2 71.0 6749 3 US-08-796-101-38 Sequence 38, Appl
249 14.2 71.0 6749 3 US-09-085-273-71 Sequence 71, Appl
250 14.2 71.0 6749 3 US-09-916-963-71 Sequence 71, Appl
251 14.2 71.0 9208 4 US-09-068-506-1 Sequence 1, Appl
252 14.2 71.0 9208 4 US-09-392-812A-5 Sequence 5, Appl
253 14.2 71.0 14758 3 US-08-975-080-35 Sequence 35, Appl
254 14.2 71.0 14796 3 US-09-630-706-10 Sequence 10, Appl
255 14.2 71.0 14796 3 US-09-396-694B-3 Sequence 3, Appl
256 14.2 71.0 33342 4 US-08-311-731A-121 Sequence 121, Appl
257 14.2 71.0 43280 2 US-08-804-227C-1 Sequence 1, Appl
258 14 70.0 489 4 US-09-252-991A-6875 Sequence 6875, Ap
259 14 70.0 624 4 US-09-252-991A-13629 Sequence 13629, A
260 14 70.0 801 4 US-09-252-991A-13819 Sequence 13819, A
261 14 70.0 885 4 US-09-489-039A-5473 Sequence 5473, Ap
262 14 70.0 1041 4 US-09-543-681A-321 Sequence 321, Ap
263 14 70.0 99916 4 US-09-816-095-3 Sequence 3, Appl
264 13.8 69.0 267 4 US-09-313-294A-6252 Sequence 6252, Ap
265 13.8 69.0 274 4 US-09-313-294A-5136 Sequence 5136, Ap
266 13.8 69.0 276 4 US-09-313-294A-5375 Sequence 5375, Ap
267 13.8 69.0 294 4 US-09-313-294A-5319 Sequence 5319, Ap
268 13.8 69.0 324 4 US-09-252-991A-14145 Sequence 14145, A
269 13.8 69.0 351 4 US-09-252-991A-4464 Sequence 4464, Ap
270 13.8 69.0 351 4 US-09-252-991A-12618 Sequence 12618, A
271 13.8 69.0 386 4 US-09-313-294A-4261 Sequence 4261, Ap
272 13.8 69.0 404 4 US-09-612-473-43 Sequence 43, Appl
273 13.8 69.0 435 4 US-09-252-991A-6670 Sequence 6670, Ap
274 13.8 69.0 456 4 US-09-252-991A-8984 Sequence 8984, Ap
275 13.8 69.0 473 4 US-09-621-976-14118 Sequence 14118, A
276 13.8 69.0 481 4 US-09-644-307B-5 Sequence 5, Appl
277 13.8 69.0 513 3 US-09-517-347-3 Sequence 3, Appl
278 13.8 69.0 513 3 US-09-517-347-7 Sequence 7, Appl
279 13.8 69.0 513 4 US-09-518-232A-3 Sequence 3, Appl
280 13.8 69.0 513 4 US-09-518-232A-7 Sequence 7, Appl
281 13.8 69.0 513 4 US-09-517-358A-3 Sequence 3, Appl
282 13.8 69.0 513 4 US-09-517-358A-7 Sequence 7, Appl
283 13.8 69.0 533 3 US-09-186-250-3 Sequence 3, Appl
284 13.8 69.0 533 3 US-09-186-250-7 Sequence 7, Appl
285 13.8 69.0 612 4 US-09-252-991A-14768 Sequence 14768, A
286 13.8 69.0 624 4 US-09-134-001C-36 Sequence 36, Appl
287 13.8 69.0 627 4 US-09-517-358A-1 Sequence 1, Appl
288 13.8 69.0 642 3 US-09-517-347-1 Sequence 1, Appl
289 13.8 69.0 642 3 US-09-517-347-5 Sequence 5, Appl
290 13.8 69.0 642 4 US-09-518-232A-1 Sequence 1, Appl
291 13.8 69.0 642 4 US-09-518-232A-5 Sequence 5, Appl
292 13.8 69.0 642 4 US-09-517-358A-1 Sequence 1, Appl
293 13.8 69.0 642 4 US-09-517-358A-5 Sequence 5, Appl
294 13.8 69.0 652 3 US-09-186-250-5 Sequence 5, Appl
295 13.8 69.0 672 4 US-09-252-991A-13019 Sequence 13019, A
296 13.8 69.0 684 3 US-08-998-416-754 Sequence 754, App
297 13.8 69.0 696 4 US-09-408-020-17 Sequence 17, Appl
298 13.8 69.0 703 3 US-09-313-300-6 Sequence 6, Appl
299 13.8 69.0 738 4 US-09-252-991A-13548 Sequence 13548, A
300 13.8 69.0 738 4 US-09-252-991A-13548 Sequence 13548, A

ALIGNMENTS

RESULT 1
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24

US-09-657-042A-3
; Sequence 3, Application US/09657042A
; Patent No. 6329203
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-1 EXPRESSION
; FILE REFERENCE: RTS-0148
; CURRENT APPLICATION NUMBER: US/09/657,042A
; CURRENT FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (79) ... (3399)
US-09-657-042A-3

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2 4403765
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 100.0%; Score 20; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCACCCTCGACTT 20
DB 3233708 GCGCGCACCCTCGACTT 3233727

RESULT 2

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 100.0%; Score 20; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCACCCTCGACTT 20
DB 3239432 GCGCGCACCCTCGACTT 3239451

RESULT 3

US-09-657-042A-3
; Sequence 3, Application US/09657042A
; Patent No. 6329203
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-1 EXPRESSION
; FILE REFERENCE: RTS-0148
; CURRENT APPLICATION NUMBER: US/09/657,042A
; CURRENT FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (79) ... (3399)
US-09-657-042A-3

Query Match	Best Local Similarity	Score	DB 4	Length	Indels	Mismatches	Gaps
Query Match	Best Local Similarity	82.0%;	Score 16.4;	DB 4; Length 3600;			
Matches 17;	Conservative 0;				1;	Indels 0;	Gaps 0;
QY	3	CGCGCACCACTCGACTT	20				
DB	3034	CGAGCACCACTCGACTT	3051				
RESULT 4	US-09-252-991A-7285						
	Sequence 7285, Application US/09252991A						
	Patent No. 6551795						
	GENERAL INFORMATION:						
	APPLICANT: Marc J. Rubenfield et al.						
	TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS						
	FILE REFERENCE: 107196.136						
	CURRENT APPLICATION NUMBER: US/09/252,991A						
	CURRENT FILING DATE: 1999-02-18						
	PRIOR APPLICATION NUMBER: US 60/074,788						
	PRIOR FILING DATE: 1998-02-18						
	PRIOR APPLICATION NUMBER: US 60/094,190						
	PRIOR FILING DATE: 1998-07-27						
	NUMBER OF SEQ ID NOS: 33142						
	SEQ ID NO 7285						
	LENGTH: 987						
	TYPE: DNA						
	ORGANISM: Pseudomonas aeruginosa						
	US-09-252-991A-7285						
Query Match	Best Local Similarity	79.0%;	Score 15.8;	DB 4; Length 987;			
Matches 17;	Conservative 0;				2;	Indels 0;	Gaps 0;
QY	1	CGCGCGCACCACTCGACT	19				
DB	134	CGCGCGCACCACTCGGCT	152				
RESULT 5	US-09-489-039A-2765/c						
	Sequence 2765, Application US/09489039A						
	Patent No. 6610836						
	GENERAL INFORMATION:						
	APPLICANT: Gary Berton et. al						
	TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA						
	FILE REFERENCE: 2709.2004001						
	CURRENT APPLICATION NUMBER: US/09/489,039A						
	CURRENT FILING DATE: 2000-01-27						
	PRIOR APPLICATION NUMBER: US 60/117,747						
	PRIOR FILING DATE: 1999-01-23						
	NUMBER OF SEQ ID NOS: 14342						
	SEQ ID NO 2765						
	LENGTH: 1083						
	TYPE: DNA						
	ORGANISM: Klebsiella pneumoniae						
	US-09-489-039A-2765						
Query Match	Best Local Similarity	79.0%;	Score 15.8;	DB 4; Length 1083;			
Matches 17;	Conservative 0;				2;	Indels 0;	Gaps 0;
QY	2	CGCGCGCACCACTCGACTT	20				
DB	683	CGCGCGCACCACTCCACCT	665				
RESULT 6	US-09-252-991A-7174						
	Sequence 7174, Application US/09252991A						

```

; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 114
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(113)
US-09-536-977-37

```

Query Match	77.0%	Score 15.4;	DB 4;	Length 114;
Best Local Similarity	94.1%;	Pred. No. 2e-02;		
Matches 16;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps	0;			

Qy 1 GCGCGCACCACTCGA 17
|||
Db 77 GCGCTGCACCACTCGA 61

RESULT 9

US-09-536-977-59/c
; Sequence 59, Application US/09536977
; Patent No. 6649409

; PACELIC NO. 6643403
; GENERAL INFORMATION:

```

1  GENERAL INFORMATION:
2  APPLICANT: FONSGBAARD, ANDERS
3  TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
4  TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
5  TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
6  TITLE OF INVENTION: EX08 CONSTRUCTS
7  FILE REFERENCE: 030307/0169
8  CURRENT APPLICATION NUMBER: US/09/536,977
9  CURRENT FILING DATE: 2000-03-29
10 PRIOR APPLICATION NUMBER: 60/128,558
11 PRIOR FILING DATE: 1999-04-09
12 PRIOR APPLICATION NUMBER: DK PA1999 00427
13 PRIOR FILING DATE: 1999-03-29
14 NUMBER OF SEQ ID NOS: 118
15 SOFTWARE: PatentIn Ver. 2.1

```

: LENGTH: 272

```

? LENGTH: 272
? TYPE: DNA
? ORGANISM: Human immunodeficiency virus
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (3)..[263]
?

```

US-09-536-977-59

Query Match 77.0%; Score 15.4; DB 4; Length 272;
Best Local Similarity 94.1%; Pred. No. 2.1e+02;
Matches 16: Conservative 0; Mismatches 1; Indels

Qy 1 GCGCGCACCACTCGA 17
pb 200 GCGCTGCACCACTCGA 184

RESULT 10

US-09-252-991A-12478/C
; Sequence 12478, Application US/09252991A
; Patent No. 6551795

; FALCIC NO. 6531735
: GENERAL INFORMATION:

? APPLICANT: MARCO RUBENFELD ET AL.
 ? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ? TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ? FILE REFERENCE: 107196.136
 ? CURRENT APPLICATION NUMBER: US/09/252,991A
 ? CURRENT FILING DATE: 1999-02-18
 ? PRIOR APPLICATION NUMBER: US 60/074,788
 ? PRIOR FILING DATE: 1998-02-18

```

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12478
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-250-991A-12478

```

Query Match 77.0%; Score 15.4; DB 4; Length 453;
Best Local Similarity 94.1%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels

QY
1 GCGCGCACCACTCGA 17

DB
338 GCGCTGCACCACTCGA 322

RESULT 11

US-09-252-991A-12406
; Sequence 12406, Application US/09252991A
; Patent No. 6551795

: GENERAL INFORMATION:

```

?
? APPLICANT: Marc J. Rubenfield et al.
? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
? TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
? FILE REFERENCE: 107196.136
? CURRENT APPLICATION NUMBER: US/09/252,991A
? CURRENT FILING DATE: 1999-02-18
? PRIOR APPLICATION NUMBER: US 60/074,788
? PRIOR FILING DATE: 1998-02-18
? PRIOR APPLICATION NUMBER: US 60/094,190
? PRIOR FILING DATE: 1998-07-27
? NUMBER OF SEQ ID NOS: 33142
? SEQ ID NO 12406
? LENGTH: 495
?

```

Query Match

Query Match 77.0%; Score 15.4; DB 4; Length 495;
Best Local Similarity 94.1%; Pred. No. 2.1e+02;
Matches 16: Conservative 0; Mismatches 1; Indels

Qy 2 CGCCGACCACTCGAC 18
pB 393 CGCCGGCCACCTCGAC 409

RESULT 12

US-09-252-991A-12111/c
; Sequence 12111, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

```

: APPLICANT: Marc J. Rubenfield et al.
:
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
:
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
:
: FILE REFERENCES: 107196.136
:
: CURRENT APPLICATION NUMBER: US/09/252,991A
:
: CURRENT FILING DATE: 1999-02-18
:
: PRIOR APPLICATION NUMBER: US 60/074,788
:
: PRIOR FILING DATE: 1998-02-18
:
: PRIOR APPLICATION NUMBER: US 60/094,190
:
: PRIOR FILING DATE: 1998-07-27
:
: NUMBER OF SEQ ID NOS: 33142

```

Query Match

Query Match	77.0%	Score 15.4;	DB 4;	Length 525;
Best Local Similarity	94.1%	Pred. No. 2.1e+02;		

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CGCGCACCACTCGAC 18
Db 521 CGCGCGCACCACTCGAC 505

RESULT 13

US-09-252-991A-12360
; Sequence 12360, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12360
; LENGTH: 690
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12360

Query Match 77.0%; Score 15.4; DB 4; Length 690;
Best Local Similarity 94.1%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CGCGCACCACTCGAC 18
Db 406 CGCGCGCACCACTCGAC 422

RESULT 14

US-09-252-991A-12250/c
; Sequence 12250, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12250
; LENGTH: 924
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12250

Query Match 77.0%; Score 15.4; DB 4; Length 924;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CGCGCACCACTCGAC 18
Db 534 CGCGCGCACCACTCGAC 518

RESULT 15

US-09-252-991A-12311
; Sequence 12311, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12311
; LENGTH: 1011
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12311

Query Match 77.0%; Score 15.4; DB 4; Length 1011;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CGCGCACCACTCGAC 18
Db 20 CGCGCGCACCACTCGAC 36

RESULT 16

US-09-536-977-75/c
; Sequence 75, Application US/09536977
; Patent No. 6649409

; GENERAL INFORMATION:

; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; TITLE OF INVENTION: BX08 CONSTRUCTS
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536,977
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128,558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 75
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1035)
US-09-536-977-75

Query Match 77.0%; Score 15.4; DB 4; Length 1038;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGCGCGCACCACTCGA 17
Db 972 CGCGCGCACCACTCGA 956

RESULT 17

US-09-522-714-15
; Sequence 15, Application US/09522714
; Patent No. 6563020

; GENERAL INFORMATION:

; APPLICANT: Simmons, Carl R.
; APPLICANT: Yalpani, Nasser
; TITLE OF INVENTION: Maize Chitinases and Their Use in
; TITLE OF INVENTION: Enhancing Disease Resistance in Crop Plants

```
; FILE REFERENCE: 1100
; CURRENT APPLICATION NUMBER: US/09/522.714
; CURRENT FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 60/125.915
; EARLIER FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 1085
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46)...(966)
US-09-522-714-15

Query Match      77.0%; Score 15.4; DB 4; Length 1085;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCGCACCACTCGAC 18
    |||||
Db 548 CGCGCGCACCACTCGAC 564

RESULT 18
US-09-252-991A-12185/c
; Sequence 12185, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12185
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12185

Query Match      77.0%; Score 15.4; DB 4; Length 1143;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCGCACCACTCGAC 18
    |||||
Db 43 CGCGCGCACCACTCGAC 27

RESULT 19
US-09-252-991A-13158
; Sequence 13158, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13158

; FILE REFERENCE: 1404
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13158

Query Match      77.0%; Score 15.4; DB 4; Length 1404;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCACTCGA 17
    |||||
Db 679 GCGCTGCACCACTCGA 695

RESULT 20
US-09-252-991A-15435
; Sequence 15435, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15435
; LENGTH: 2163
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1464)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-15435

Query Match      77.0%; Score 15.4; DB 4; Length 2163;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGCGCGCACCACTCGAC 18
    |||||
Db 1175 CGCGCGCACCACTCGAC 1191

RESULT 21
US-09-536-977-71/c
; Sequence 71, Application US/09536977
; Patent No. 6649409
; GENERAL INFORMATION:
; APPLICANT: FOMSGAARD, ANDERS
; TITLE OF INVENTION: METHOD FOR PRODUCING A NUCLEOTIDE SEQUENCE CONSTRUCT
; TITLE OF INVENTION: WITH OPTIMIZED CODONS FOR AN HIV GENETIC VACCINE BASED
; TITLE OF INVENTION: ON A PRIMARY, EARLY HIV ISOLATE AND SYNTHETIC ENVELOPE
; TITLE OF INVENTION: BX08 CONSTRUCTS
; FILE REFERENCE: 030307/0169
; CURRENT APPLICATION NUMBER: US/09/536.977
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/128.558
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: DK PA1999 00427
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 71
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Human immunodeficiency virus
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (1)...(2460)
US-09-536-977-71

Query Match      77.0%; Score 15.4; DB 4; Length 2469;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCGCGGCACCACTCGA 17
Db      2397 GCGCTGCACCACTCGA 2381
|||||

RESULT 22
US-09-252-991A-12788/c
; Sequence 12788, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12788
; LENGTH: 2844
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12788

Query Match      77.0%; Score 15.4; DB 4; Length 2844;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCGCGGCACCACTCGA 17
Db      2184 GCGCTGCACCACTCGA 2168
|||||

RESULT 23
US-09-252-991A-12644/c
; Sequence 12644, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12644
; LENGTH: 3342
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12644

Query Match      77.0%; Score 15.4; DB 4; Length 3342;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCGCGGCACCACTCGA 17
Db      1690 GCGCTGCACCACTCGA 1674
|||||

; NAME/KEY: CDS
; LOCATION: (1)...(2460)
US-09-536-977-71

Query Match      77.0%; Score 15.4; DB 4; Length 2469;
Best Local Similarity 94.1%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCGCGGCACCACTCGA 17
Db      2397 GCGCTGCACCACTCGA 2381
|||||

RESULT 24
US-09-595-684B-38/c
; Sequence 38, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 6972
; TYPE: DNA
; ORGANISM: Human
US-09-595-684B-38

Query Match      77.0%; Score 15.4; DB 4; Length 6972;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCGCGGCACCACTCGA 17
Db      4866 GCGCGGCACCACTCGA 4850
|||||

RESULT 25
US-09-268-163-9
; Sequence 9, Application US/09268163B
; Patent No. 6353091
; GENERAL INFORMATION:
; APPLICANT: Lipscombe, Diare
; APPLICANT: Schorge, Stephanie
; TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF
; FILE REFERENCE: B1055/7000
; CURRENT APPLICATION NUMBER: US/09/268,163B
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 60/077,901
; EARLIER FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 7011
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..7008
US-09-268-163-9

Query Match      77.0%; Score 15.4; DB 4; Length 7011;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCGCGGCACCACTCGA 17
Db      5164 GCGCGGCACCACTCGA 5180
|||||

RESULT 26
US-09-620-312D-1083/c
; Sequence 1083, Application US/09620312D
```


Patent No. 6569662
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Zhao, Qing A.
APPLICANT: Wehrman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Duanui
APPLICANT: Wang, Zhiwei
APPLICANT: John Tillinghast
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2B
CURRENT APPLICATION NUMBER: US/09/620,312D
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1105
SOFTWARE: pc_Fl_genes Version 1.0
SEQ ID NO 1083
LENGTH: 8309
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (162)..(2843)
US-09-620-312D-1083

Query Match 77.0%; Score 15.4; DB 4; Length 8309;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCACCACCTCGA 17
DB 4967 GCGCGCACCACCA 4951

RESULT 27

US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 77.0%; Score 15.4; DB 3; Length 4403765;
Best Local Similarity 94.1%; Pred. No. 1.1e-02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGCACCACCTCGAC 18
DB 1272177 GCGCGCAGCACCTCGAC 1272161

RESULT 28

US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 77.0%; Score 15.4; DB 3; Length 4411529;
Best Local Similarity 94.1%; Pred. No. 1.1e-02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCGCGCACCACCTCGAC 18
DB 1272707 GCGCGCAGCACCTCGAC 1272691

RESULT 29

US-09-621-976-1316/c
Sequence 1316, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 1316
LENGTH: 333
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 13..333
US-09-621-976-1316

Query Match 76.0%; Score 15.2; DB 4; Length 333;
Best Local Similarity 85.0%; Pred. No. 2.6e-02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGCGCACCACCTCGACTT 20
DB 285 GAGCGCACCACCACTCGGCTT 266

RESULT 30
US-09-252-991A-8968
; Sequence 9968, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8968
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8968

Query Match 76.0%; Score 15.2; DB 4; Length 588;
Best Local Similarity 85.0%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GGCGCGCACCACTCGACTT 20
Db 102 GAGCGCGCACCACTCGCGTT 121

Search completed: June 20, 2004, 11:46:10
Job time : 37.2414 secs

C 88	15.2	76.0	1173	16	US-10-369-493-31555	Sequence 31555, A	161	14.8	74.0	1038	16	US-10-369-493-39798	Sequence 39798, A
C 89	15.2	76.0	1176	16	US-10-369-493-43218	Sequence 43218, A	162	14.8	74.0	1050	16	US-10-369-493-43301	Sequence 43301, A
C 90	15.2	76.0	1238	13	US-10-425-114-26641	Sequence 26641, A	C 163	14.8	74.0	1055	16	US-10-369-493-34346	Sequence 34346, A
C 91	15.2	76.0	1260	13	US-10-422-366-3	Sequence 3, Appli	C 164	14.8	74.0	1059	15	US-10-156-761-3160	Sequence 3160, Ap
C 92	15.2	76.0	1260	13	US-10-422-366-5	Sequence 5, Appli	C 165	14.8	74.0	1074	16	US-10-369-493-31921	Sequence 31921, A
C 93	15.2	76.0	1287	13	US-10-425-114-1563	Sequence 1563, Ap	C 166	14.8	74.0	1188	13	US-10-282-122A-31591	Sequence 31591, A
C 94	15.2	76.0	1306	16	US-10-369-493-43157	Sequence 43157, A	C 167	14.8	74.0	1254	15	US-10-214-446-25	Sequence 25, Appli
C 95	15.2	76.0	1317	16	US-10-028-445-4	Sequence 4, Appli	C 168	14.8	74.0	1380	13	US-10-417-700A-134	Sequence 134, App
C 96	15.2	76.0	1428	16	US-10-369-493-35927	Sequence 35927, A	C 169	14.8	74.0	1429	13	US-10-425-114-15471	Sequence 15471, A
C 97	15.2	76.0	1446	16	US-10-369-493-36028	Sequence 36028, A	C 170	14.8	74.0	1440	13	US-10-417-700A-70	Sequence 70, Appli
C 98	15.2	76.0	1548	13	US-10-156-761-5601	Sequence 5601, Ap	C 171	14.8	74.0	1489	13	US-10-425-114-14295	Sequence 14295, A
C 99	15.2	76.0	1590	13	US-10-282-122A-31781	Sequence 31781, A	C 172	14.8	74.0	1556	15	US-10-081-872-121	Sequence 121, App
C 100	15.2	76.0	1722	15	US-10-156-761-4387	Sequence 4387, Ap	C 173	14.8	74.0	1556	16	US-10-385-305-121	Sequence 121, App
C 101	15.2	76.0	1826	13	US-10-028-445-1	Sequence 1, Appli	C 174	14.8	74.0	1560	15	US-10-156-761-5073	Sequence 5073, Ap
C 102	15.2	76.0	1959	13	US-10-282-122A-25712	Sequence 25712, A	C 175	14.8	74.0	1599	15	US-10-156-761-3074	Sequence 3074, Ap
C 103	15.2	76.0	1960	13	US-10-425-114-14258	Sequence 14258, A	C 176	14.8	74.0	1622	16	US-10-260-238-1073	Sequence 1073, Ap
C 104	15.2	76.0	2058	16	US-10-369-493-43117	Sequence 43117, A	C 177	14.8	74.0	1665	16	US-10-369-493-43323	Sequence 43323, A
C 105	15.2	76.0	2244	16	US-10-320-797-2288	Sequence 2288, Ap	C 178	14.8	74.0	1701	13	US-10-282-122A-32068	Sequence 32068, A
C 106	15.2	76.0	2289	13	US-10-282-122A-39166	Sequence 39166, Ap	C 179	14.8	74.0	1714	13	US-10-425-114-1753	Sequence 1753, Ap
C 107	15.2	76.0	2649	16	US-10-320-797-1288	Sequence 1288, Ap	C 180	14.8	74.0	1731	13	US-10-282-122A-26348	Sequence 26348, A
C 108	15.2	76.0	2894	13	US-10-282-122A-33613	Sequence 33613, A	C 181	14.8	74.0	1734	13	US-10-282-122A-28433	Sequence 28433, A
C 109	15.2	76.0	2933	10	US-09-809-391-165	Sequence 165, App	C 182	14.8	74.0	1790	13	US-10-425-114-3920	Sequence 3920, Ap
C 110	15.2	76.0	2933	13	US-09-882-171-165	Sequence 165, App	C 183	14.8	74.0	1857	13	US-10-282-122A-23284	Sequence 23284, A
C 111	15.2	76.0	3233	13	US-10-164-861-165	Sequence 165, App	C 184	14.8	74.0	2000	16	US-10-260-238-1672	Sequence 1672, Ap
C 112	15.2	76.0	3195	9	US-09-922-217-1114	Sequence 1114, Ap	C 185	14.8	74.0	2000	16	US-10-260-238-2619	Sequence 2619, Ap
C 113	15.2	76.0	3195	14	US-10-025-380-1114	Sequence 1114, Ap	C 186	14.8	74.0	2067	15	US-10-156-761-6391	Sequence 6391, Ap
C 114	15.2	76.0	3198	16	US-10-314-434-98	Sequence 98, Appli	C 187	14.8	74.0	2148	15	US-10-156-761-3210	Sequence 3210, Ap
C 115	15.2	76.0	3276	10	US-09-809-391-298	Sequence 298, App	C 188	14.8	74.0	2162	13	US-10-425-114-25098	Sequence 25098, A
C 116	15.2	76.0	3276	10	US-09-882-171-298	Sequence 298, App	C 189	14.8	74.0	2175	10	US-09-764-891-7546	Sequence 7546, Ap
C 117	15.2	76.0	3276	13	US-10-164-861-639	Sequence 639, App	C 190	14.8	74.0	2247	13	US-10-282-122A-25458	Sequence 25458, A
C 118	15.2	76.0	3776	16	US-10-264-049-639	Sequence 639, App	C 191	14.8	74.0	2248	13	US-10-425-114-1870	Sequence 1870, A
C 119	15.2	76.0	4649	9	US-10-320-797-288	Sequence 288, App	C 192	14.8	74.0	2423	15	US-10-156-761-3443	Sequence 3443, Ap
C 120	15.2	76.0	8391	16	US-09-870-759-113	Sequence 113, App	C 193	14.8	74.0	2463	13	US-10-282-122A-13810	Sequence 13810, A
C 121	15.2	76.0	8391	10	US-09-751-708A-113	Sequence 113, App	C 194	14.8	74.0	2463	16	US-10-369-493-44476	Sequence 44476, A
C 122	15.2	76.0	9098	13	US-10-221-278-34	Sequence 34, Appli	C 195	14.8	74.0	2466	13	US-10-282-122A-11853	Sequence 11853, A
C 123	15.2	76.0	9098	15	US-10-119-428-37	Sequence 37, Appli	C 196	14.8	74.0	2466	13	US-10-369-493-44476	Sequence 44476, A
C 124	15.2	76.0	9098	16	US-10-291-172-34	Sequence 34, Appli	C 197	14.8	74.0	2511	13	US-10-282-122A-27874	Sequence 27874, A
C 125	15.2	76.0	9210	9	US-09-712-363-100	Sequence 100, App	C 198	14.8	74.0	2581	15	US-10-156-761-5779	Sequence 5779, Ap
C 126	15.2	76.0	9590	13	US-10-282-122A-33351	Sequence 3351, A	C 199	14.8	74.0	3083	16	US-10-094-749-1096	Sequence 1096, Ap
C 127	15.2	76.0	30042	15	US-10-080-170-650	Sequence 650, App	C 200	14.8	74.0	3190	16	US-10-260-238-1050	Sequence 1050, Ap
C 128	15.2	76.0	135638	15	US-10-314-657-1	Sequence 1, Appli	C 201	14.8	74.0	3668	10	US-09-927-827-21	Sequence 21, Appli
C 129	15.2	76.0	135638	15	US-10-314-657-1	Sequence 1, Appli	C 202	14.8	74.0	3668	16	US-10-084-846A-49	Sequence 49, Appli
C 130	15.2	76.0	536165	10	US-09-939-964-1	Sequence 1, Appli	C 203	14.8	74.0	5110	9	US-09-764-846-310	Sequence 310, App
C 131	15.2	76.0	9025608	13	US-10-156-761-1	Sequence 1, Appli	C 204	14.8	74.0	5110	15	US-10-091-483-310	Sequence 310, App
C 132	15	75.0	630	13	US-10-282-122A-14409	Sequence 14409, A	C 205	14.8	74.0	5489	15	US-10-311-455-2094	Sequence 2094, Ap
C 133	15	75.0	1608	15	US-10-156-761-555	Sequence 555, App	C 206	14.8	74.0	7155	15	US-10-329-079-14	Sequence 14, Appli
C 134	15	75.0	1947	13	US-10-425-114-20740	Sequence 20740, A	C 207	14.8	74.0	7155	9	US-09-764-877-2277	Sequence 2277, Ap
C 135	14.8	74.0	182	15	US-10-131-002-29	Sequence 29, Appli	C 208	14.8	74.0	9970	16	US-10-242-515-2277	Sequence 2277, Ap
C 136	14.8	74.0	182	15	US-10-084-843-29	Sequence 29, Appli	C 209	14.8	74.0	20922	13	US-09-942-025-14	Sequence 14, Appli
C 137	14.8	74.0	396	13	US-10-282-122A-25978	Sequence 25978, A	C 210	14.8	74.0	25085	15	US-10-132-134-35	Sequence 35, Appli
C 138	14.8	74.0	4923	13	US-10-282-122A-22993	Sequence 22993, A	C 211	14.8	74.0	37360	15	US-10-329-079-6	Sequence 6, Appli
C 139	14.8	74.0	525	13	US-10-425-114-5650	Sequence 5650, Ap	C 212	14.8	74.0	50543	15	US-10-132-134-25	Sequence 25, Appli
C 140	14.8	74.0	535	13	US-09-770-152-771	Sequence 771, App	C 213	14.8	74.0	59816	16	US-10-084-846A-1	Sequence 1, Appli
C 141	14.8	74.0	596	13	US-10-425-114-22439	Sequence 22439, A	C 214	14.8	74.0	59816	16	US-10-084-846A-2	Sequence 2, Appli
C 142	14.8	74.0	640	15	US-10-260-238-2784	Sequence 2784, Ap	C 215	14.8	74.0	63686	13	US-10-087-192-466	Sequence 466, App
C 143	14.8	74.0	651	15	US-10-156-761-3835	Sequence 3835, Ap	C 216	14.8	74.0	67311	13	US-09-942-025-1	Sequence 1, Appli
C 144	14.8	74.0	656	15	US-10-029-386-20822	Sequence 20822, A	C 217	14.8	74.0	68750	14	US-10-014-717-1	Sequence 1, Appli
C 145	14.8	74.0	677	13	US-10-425-114-31232	Sequence 31232, A	C 218	14.8	74.0	71989	13	US-09-727-889-2	Sequence 2, Appli
C 146	14.8	74.0	713	13	US-10-037-632-27651	Sequence 27651, A	C 219	14.8	74.0	207433	16	US-10-277-216-5	Sequence 5, Appli
C 147	14.8	74.0	713	16	US-10-037-632-27651	Sequence 27651, A	C 220	14.8	74.0	207433	17	US-10-126-022-5	Sequence 5, Appli
C 148	14.8	74.0	737	13	US-10-425-114-22763	Sequence 22763, A	C 221	14.8	72.0	45	15	US-10-238-129-14	Sequence 14, Appli
C 149	14.8	74.0	752	13	US-10-037-632-135346	Sequence 135346, A	C 222	14.4	72.0	45	15	US-10-238-667-14	Sequence 14, Appli
C 150	14.8	74.0	752	13	US-10-027-632-135347	Sequence 135347, A	C 223	14.4	72.0	135	13	US-10-085-783A-20971	Sequence 20971, A
C 151	14.8	74.0	752	16	US-10-027-632-135346	Sequence 135346, A	C 224	14.4	72.0	135	16	US-10-242-535A-20971	Sequence 20971, A
C 152	14.8	74.0	752	16	US-10-027-632-135347	Sequence 135347, A	C 225	14.4	72.0	165	13	US-10-085-783A-15623	Sequence 15623, A
C 153	14.8	74.0	864	15	US-10-156-761-3578	Sequence 3578, Ap	C 226	14.4	72.0	165	16	US-10-242-535A-10151	Sequence 10151, A
C 154	14.8	74.0	894	13	US-10-282-122A-11787	Sequence 11787, A	C 227	14.4	72.0	177	13	US-10-085-783A-10151	Sequence 10151, A
C 155	14.8	74.0	939	13	US-10-282-122A-14724	Sequence 14724, A	C 228	14.4	72.0	177	16	US-10-242-535A-10151	Sequence 10151, A
C 156	14.8	74.0	968	16	US-10-369-493-44335	Sequence 44335, A	C 229	14.4	72.0	246	16	US-10-085-783A-44778	Sequence 44778, A
C 157	14.8	74.0	993	16	US-10-084-846A-67	Sequence 67, Appli	C 230	14.4	72.0	246	16	US-10-242-535A-14778	Sequence 14778, A
C 158	14.8	74.0	1008	15	US-10-156-761-1973	Sequence 1973, Ap	C 231	14.4	72.0	269	16	US-10-264-049-1689	Sequence 1689, Ap
C 159	14.8	74.0	1035	16	US-10-369-493-39047	Sequence 39047, A	C 232	14.4	72.0	269	16	US-10-085-783A-32925	Sequence 32925, A
C 160	14.8	74.0	1038	16	US-10-369-493-39414	Sequence 39414, A	C 233	14.4	72.0	264	16	US-10-242-535A-32925	Sequence 32925, A

C 234	14.4	72.0	271	13	US-10-085-783A-31525	Sequence 31525, A
C 235	14.4	72.0	271	16	US-10-242-535A-31525	Sequence 31525, A
C 236	14.4	72.0	307	13	US-09-852-385-2	Sequence 2, Appl
C 237	14.4	72.0	308	9	US-09-960-352-8755	Sequence 8755, Ap
C 238	14.4	72.0	326	9	US-09-960-352-9664	Sequence 9664, Ap
C 239	14.4	72.0	331	14	US-10-032-159A-24	Sequence 24, Appl
C 240	14.4	72.0	336	13	US-10-085-783A-3213	Sequence 3213, A
C 241	14.4	72.0	336	16	US-10-242-535A-3213	Sequence 3213, A
C 242	14.4	72.0	341	9	US-09-960-352-10028	Sequence 10028, A
C 243	14.4	72.0	358	16	US-10-264-049-1766	Sequence 1766, Ap
C 244	14.4	72.0	363	13	US-10-085-783A-23016	Sequence 23016, A
C 245	14.4	72.0	363	16	US-10-242-535A-23016	Sequence 23016, A
C 246	14.4	72.0	373	13	US-10-085-783A-16568	Sequence 16568, A
C 247	14.4	72.0	373	13	US-10-085-783A-45716	Sequence 45716, A
C 248	14.4	72.0	373	16	US-10-242-535A-45716	Sequence 45716, A
C 249	14.4	72.0	374	13	US-10-085-783A-52065	Sequence 52065, A
C 250	14.4	72.0	374	16	US-10-242-535A-52065	Sequence 52065, A
C 251	14.4	72.0	390	10	US-09-971-393-234	Sequence 234, Appl
C 252	14.4	72.0	390	13	US-10-296-115-583	Sequence 583, Appl
C 253	14.4	72.0	405	13	US-10-424-599-21175	Sequence 21175, A
C 254	14.4	72.0	428	13	US-10-156-761-1554	Sequence 1554, Ap
C 255	14.4	72.0	453	13	US-10-264-049-657	Sequence 657, Appl
C 256	14.4	72.0	460	16	US-09-918-995-1033	Sequence 1033, Ap
C 257	14.4	72.0	472	10	US-10-369-493-40619	Sequence 40619, A
C 258	14.4	72.0	480	16	US-09-918-995-19184	Sequence 19184, A
C 259	14.4	72.0	485	10	US-09-918-995-19750	Sequence 19750, A
C 260	14.4	72.0	488	10	US-09-918-995-2212	Sequence 2212, Ap
C 261	14.4	72.0	508	10	US-09-918-995-350	Sequence 350, Appl
C 262	14.4	72.0	524	9	US-09-925-299-350	Sequence 350, Appl
C 263	14.4	72.0	524	10	US-10-156-761-5927	Sequence 5927, Ap
C 264	14.4	72.0	624	15	US-10-156-761-5927	Sequence 5927, Ap
C 265	14.4	72.0	630	13	US-10-425-114-24043	Sequence 24043, A
C 266	14.4	72.0	633	15	US-10-156-761-1655	Sequence 1655, Ap
C 267	14.4	72.0	670	15	US-10-132-350-29	Sequence 29, Appl
C 268	14.4	72.0	670	13	US-10-132-350-31	Sequence 31, Appl
C 269	14.4	72.0	678	13	US-10-424-599-13285	Sequence 13285, A
C 270	14.4	72.0	683	13	US-10-424-599-129203	Sequence 129203, A
C 271	14.4	72.0	815	16	US-10-074-978A-23	Sequence 23, Appl
C 272	14.4	72.0	856	13	US-10-027-632-161018	Sequence 161018, A
C 273	14.4	72.0	856	13	US-10-027-632-161019	Sequence 161019, A
C 274	14.4	72.0	856	16	US-10-027-632-161018	Sequence 161018, A
C 275	14.4	72.0	856	16	US-10-027-632-161019	Sequence 161019, A
C 276	14.4	72.0	862	16	US-10-260-238-1324	Sequence 1324, Ap
C 277	14.4	72.0	870	15	US-10-156-761-3118	Sequence 3118, Ap
C 278	14.4	72.0	946	13	US-10-425-114-33935	Sequence 33935, A
C 279	14.4	72.0	977	15	US-10-029-386-24193	Sequence 24193, A
C 280	14.4	72.0	1002	15	US-10-156-761-4347	Sequence 4347, Ap
C 281	14.4	72.0	1088	13	US-10-425-114-2306	Sequence 2306, Ap
C 282	14.4	72.0	1094	16	US-10-158-057-27	Sequence 27, Appl
C 283	14.4	72.0	1101	13	US-10-382-122A-25531	Sequence 25531, A
C 284	14.4	72.0	1113	13	US-10-336-472-73	Sequence 73, Appl
C 285	14.4	72.0	1113	13	US-10-336-472-77	Sequence 77, Appl
C 286	14.4	72.0	1113	13	US-10-336-472-83	Sequence 83, Appl
C 287	14.4	72.0	1116	15	US-10-156-761-5828	Sequence 5828, Ap
C 288	14.4	72.0	1149	13	US-10-412-699B-991	Sequence 991, Appl
C 289	14.4	72.0	1149	16	US-10-374-780A-487	Sequence 487, Appl
C 290	14.4	72.0	1197	15	US-10-260-715-11	Sequence 11, Appl
C 291	14.4	72.0	1200	15	US-10-146-733-3	Sequence 3, Appl
C 292	14.4	72.0	1200	15	US-10-146-733-6	Sequence 6, Appl
C 293	14.4	72.0	1216	9	US-09-746-491-21	Sequence 21, Appl
C 294	14.4	72.0	1216	13	US-10-336-472-75	Sequence 75, Appl
C 295	14.4	72.0	1233	13	US-10-282-122A-23596	Sequence 23596, A
C 296	14.4	72.0	1286	9	US-09-746-491-56	Sequence 56, Appl
C 297	14.4	72.0	1286	9	US-09-746-491-57	Sequence 57, Appl
C 298	14.4	72.0	1286	16	US-10-173-999-41	Sequence 41, Appl
C 299	14.4	72.0	1389	15	US-10-156-761-3920	Sequence 3920, Ap
C 300	14.4	72.0	1395	16	US-10-369-493-44291	Sequence 44291, A

ALIGNMENTS

RESULT 1

US-10-282-122A-26184/c
; Sequence 26184, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26184
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26184

Query Match 100.0%; Score 20; DB 13; Length 867;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
Db 36 GCGCGCACCACTCGACTT 17
|||||

RESULT 2
US-10-282-122A-28622/c
; Sequence 28622, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert

```

; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28622
; LENGTH: 870
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28622

Query Match          100.0%; Score 20; DB 13; Length 870;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGCGGCACCACTCGACTT 20
DB 36 GCGCGGCACCACTCGACTT 17

RESULT 3
US-10-282-122A-26037/c
; Sequence 26037, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06

```

```

; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26037
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Mycobacterium avium
US-10-282-122A-26037

Query Match          92.0%; Score 18.4; DB 13; Length 849;
Best Local Similarity 95.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGGCACCACTCGACTT 20
DB 36 GCGCGGCACCACTCGACTT 17

RESULT 4
US-10-282-122A-27722/c
; Sequence 27722, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614

```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27722
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Mycobacterium leprae
US-10-282-122A-27722

Query Match      92.0%; Score 18.4; DB 13; Length 849;
Best Local Similarity 95.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GCGCGCACCACTCGACTT 20
Db      36 GCGCGCACCACTCGACTT 17

RESULT 5
US-10-282-122A-31936/c
; Sequence 31936, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Oalsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31936
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Pseudomonas putida
US-10-282-122A-31936

Query Match      85.0%; Score 17; DB 13; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 GCGCGCACCACTCGACT 19
Db      277 GCGCGCACCACTCGACT 261
```

```
RESULT 6
US-10-260-238-1112
; Sequence 1112, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 1112
; LENGTH: 846
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-260-238-1112

Query Match      84.0%; Score 16.8; DB 16; Length 846;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GCGCGCACCACTCGACTT 20
Db      646 GCGCGCACCACTCGACTT 665

RESULT 7
US-09-738-626-2272/c
; Sequence 2272, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2272
; LENGTH: 858
```

TYPE: DNA
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-2272

Query Match 84.0%; Score 16.8; DB 9; Length 858;
 Best Local Similarity 90.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 DB 36 GCGCGCACCACTCGACTT 17

RESULT 8

US-10-108-605-30
 Sequence 30, Application US/10108605
 Publication No. US20020160934A1

GENERAL INFORMATION:
 APPLICANT: Broadus, Julie
 APPLICANT: Stam, Lynn
 APPLICANT: Bachmann, Jane
 APPLICANT: Kamdar, Kim
 TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
 FILE REFERENCE: 31133B
 CURRENT APPLICATION NUMBER: US/10/108,605

PRIOR FILING DATE: 2002-03-27
 PRIOR APPLICATION NUMBER: US 09/761,142
 PRIOR FILING DATE: 2001-01-16
 PRIOR APPLICATION NUMBER: US 60/176,418
 PRIOR FILING DATE: 2000-01-14
 NUMBER OF SEQ ID NOS: 361
 SOFTWARE: PatentIn ver. 2.1
 SEQ ID NO 30
 TYPE: DNA

ORGANISM: Drosophila melanogaster
 US-10-108-605-30

Query Match 84.0%; Score 16.8; DB 14; Length 1654;
 Best Local Similarity 90.0%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 DB 239 GCTCGCACCACTCGACTT 258

RESULT 9

US-10-424-599-94765
 Sequence 94765, Application US/10424599
 Publication No. US20040031072A1

GENERAL INFORMATION:
 APPLICANT: La Rosa Thomas J
 APPLICANT: Kowalic David K
 APPLICANT: Zhou Yihua
 APPLICANT: Cao Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599

PRIOR FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 94765
 LENGTH: 1913
 TYPE: DNA

ORGANISM: Glycine max
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT3847_56584C.1
 US-10-424-599-94765

Query Match 84.0%; Score 16.8; DB 13; Length 1913;
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GCGCGCACCACTCGACTT 20
 |||||
 DB 172 GCGCGCACCACTCGACTT 191

RESULT 10

US-10-108-605-32
 Sequence 32, Application US/10108605
 Publication No. US20020160934A1

GENERAL INFORMATION:
 APPLICANT: Broadus, Julie
 APPLICANT: Stam, Lynn
 APPLICANT: Bachmann, Jane
 APPLICANT: Kamdar, Kim
 TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
 FILE REFERENCE: 31133B
 CURRENT APPLICATION NUMBER: US/10/108,605

PRIOR FILING DATE: 2002-03-27
 PRIOR APPLICATION NUMBER: US 09/761,142
 PRIOR FILING DATE: 2001-01-16
 PRIOR APPLICATION NUMBER: US 60/176,418
 PRIOR FILING DATE: 2000-01-14
 NUMBER OF SEQ ID NOS: 361
 SOFTWARE: PatentIn ver. 2.1
 SEQ ID NO 32
 LENGTH: 2163
 TYPE: DNA

ORGANISM: Drosophila melanogaster
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1)..(2163)
 OTHER INFORMATION: n = a, t, c, or g
 US-10-108-605-32

Query Match 84.0%; Score 16.8; DB 14; Length 2163;
 Best Local Similarity 90.0%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGACTT 20
 |||||
 DB 1082 GCTCGCACCACTCGACTT 1101

RESULT 11

US-09-738-626-1
 Sequence 1, Application US/09738626
 Publication No. US20020197605A1

GENERAL INFORMATION:
 APPLICANT: NAKAGAWA, SATOSHI
 APPLICANT: MIZOGUCHI, HIROSHI
 APPLICANT: ANDO, SEIKO
 APPLICANT: HAYASHI, MIKIRO
 APPLICANT: OCHIALI, KEIKO
 APPLICANT: YOKOI, HARUHIKO
 APPLICANT: TATEISHI, NAOKO
 APPLICANT: SENOH, AKIHIRO
 APPLICANT: IKEDA, MASATO
 APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738,626
 CURRENT FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/280988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0


```

; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

```

```
Query Match      84.0%; Score 16.8; DB 9; Length 3309400;
Best Local Similarity 90.0%; Pred. No. 73;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 GCGCGCACCACTCGACTT 20
||| ||| ||| ||| ||| ||| |||
Db 2189135 GCGCGCACCACTCGACTT 2189154

RESULT 12

US-10-260-238-946/c
; Sequence 946, Application US/102602038
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Eumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong

```

/ AFFILIANT: 2342, 109
/
/ TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
/
/ FILE REFERENCE: 60111-NP
/
/ CURRENT APPLICATION NUMBER: US/10/260,238
/
/ CURRENT FILING DATE: 2002-09-26
/
/ PRIOR APPLICATION NUMBER: US 60/325,448
/
/ PRIOR FILING DATE: 2001-09-26
/
/ PRIOR APPLICATION NUMBER: US 60/325,277
/
/ PRIOR FILING DATE: 2001-09-26
/
/ PRIOR APPLICATION NUMBER: US 60/370,620
/
/ PRIOR FILING DATE: 2002-04-04
/
/ NUMBER OF SEQ ID NOS: 6077
/
/ SEQ ID NO 946
/
/ LENGTH: 965
/
/ TYPE: DNA
/
/ ORGANISM: Oryza sativa
/
/ US-10-260-238-946

```

```
Query Match      82.0%; Score 16.4; DB 16; Length 965;
Best Local Similarity 94.4%; Pred. No. 2.2e+02;
Matches 17: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 GCGCGCACCACTCGAC 18
db 36 GCGCGCACCACTCGTC 19

RESULT 13

```

US/10-156-761-4971
; Sequence 4971, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: ISHIKAWA, HIROSHI
; APPLICANT: SHIEBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156.761
; CURRENT FILING DATE: 2002-05-29

```

```

; PRIOR APPLICATION NUMBER: JP 2001-2040899
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-2726977
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4971
; LENGTH: 1272

```

```

; LOCUS: 1272
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1272)
US-10-156-761-4971

```

Query Match : 82.0%; Score 16.4; DB 15; Length 1272;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0;

QY 1 GCGCGCACCACTCGAC 18
|||
Db 301 GCGCGCACCACTCGAC 318

RESULT 14

```

US-10-425-114-30185
; Sequence 30185, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecule
; TITLE OF INVENTION: Plants and Uses Thereof
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30185
; LENGTH: 1274
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-OSFLM202
US-10-425-114-30185

```

```
Query Match      82.0%; Score 16.4; DB 13; Length 1274;
Best Local Similarity 94.4%; Pred. NO. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 GCGCGCACCACTCGAC 18
Db 818 GCGCGCACCACTCTAC 835

RESULT 15

US-10-369-493-32442/c
Sequence 32442, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 32442
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Ralstonia metallidurans
US-10-369-493-32442

Query Match 82.0%; Score 16.4; DB 16; Length 1542;
Best Local Similarity 94.4%; Pred. NO. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGCGCACCACTCGAC 18
|||
DB 267 GCGCGCACCACTCGCC 250

RESULT 16

; Sequence 29, Application US/10157031
; Publication No. US20030108990A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 3587
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-157-031-29

Query Match 82.0%; Score 16.4; DB 15; Length 3587;
Best Local Similarity 94.4%; Pred. NO. 2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCGCACCACTCGACTT 20
|||
DB 3003 GCAGCACCACTCGACTT 3020

RESULT 17

US-10-380-126-3
; Sequence 3, Application US/10380126
; Publication No. US20040029824A1
; GENERAL INFORMATION:
; APPLICANT: Isis Pharmaceuticals, Inc.
; APPLICANT: C. Frank Bennett
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF GLIOMA-ASSOCIATED ONCOGENE-1 EXPRESSION
; FILE REFERENCE: RISP-0175
; CURRENT APPLICATION NUMBER: US/10/380,126
; CURRENT FILING DATE: 2003-03-10
; PRIOR APPLICATION NUMBER: 09/657,042
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 88
; SEQ ID NO 3
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (79)....(3399)
US-10-380-126-3

Query Match 82.0%; Score 16.4; DB 13; Length 3600;

Best Local Similarity 94.4%; Pred. NO. 2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCGCACCACTCGACTT 20
|||
DB 3034 GCAGCACCACTCGACTT 3051

RESULT 18

US-09-764-868-1503/c
; Sequence 1503, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1503
; LENGTH: 8522
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4173)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-1503

Query Match 82.0%; Score 16.4; DB 9; Length 8522;
Best Local Similarity 94.4%; Pred. NO. 1.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCGCACCACTCGACTT 20
|||
DB 7917 GCAGCACCACTCGACTT 7900

RESULT 19

US-09-764-868-1504/c
; Sequence 1504, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1504
; LENGTH: 14486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-1504

Query Match 82.0%; Score 16.4; DB 9; Length 14486;
Best Local Similarity 94.4%; Pred. NO. 1.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GCGCACCACTCGACTT 20
|||
DB 7921 GCAGCACCACTCGACTT 7904

RESULT 20

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIOYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Query Match 82.0%; Score 16.4; DB 15; Length 9025608;
Best Local Similarity 94.4%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GCGCGGCACACCTCGAC 18
Db 6035826 GCGCGGCACACCTCGAC 6035843
RESULT 21
US-09-294-093B-2581/c
Sequence 2581, Application US/09294093B
Patent No. US20010051335A1
GENERAL INFORMATION:
APPLICANT: Lalgudi, Raghunath, V.
APPLICANT: Ito, Laura, Y.
APPLICANT: Sherman, Bradley, K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
FILE REFERENCE: PL-0009 US
CURRENT APPLICATION NUMBER: US/09/294,093B
CURRENT FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: 60/082,567
PRIOR FILING DATE: April 21, 1998
NUMBER OF SEQ ID NOS: 6207
SOFTWARE: PERL Program
SEQ ID NO 2581
LENGTH: 293
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20010051335A1 700346329H1
NAME/KEY: unsure
LOCATION: 2, 39, 109, 111, 115, 132
OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-2581
Query Match 80.0%; Score 16; DB 9; Length 293;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGGCACACCTCG 16
Db 91 GCGCGGCACACCTCG 76
RESULT 22
US-10-425-114-4008/c
Sequence 4008, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 4008
LENGTH: 1197
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700346329_FLI
US-10-425-114-4008
Query Match 80.0%; Score 16; DB 13; Length 1197;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGGCACACCTCG 16
Db 90 GCGCGGCACACCTCG 75
RESULT 23
US-10-425-114-32590/c
Sequence 32590, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 32590
LENGTH: 2043
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: UC-ZMFLMO17007E01_FLI
US-10-425-114-32590
Query Match 80.0%; Score 16; DB 13; Length 2043;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGCGGCACACCTCG 16
Db 1196 GCGCGGCACACCTCG 1181
RESULT 24
US-09-918-995-29518/c
Sequence 29518, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 29518

; LENGTH: 446

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc_feature

; LOCATION: (1)...(446)

; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-29518

Query Match 79.0%; Score 15.8; DB 10; Length 446;
Best Local Similarity 89.5%; Pred. No. 4.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGCGCGCACCTCGACTT 20

Db 114 CGCGCGCACCTCGGCTT 96

RESULT 25

US-10-138-970A-1

; Sequence 1, Application US/10138970A

; Publication No. US20030079255A1

; GENERAL INFORMATION:

; APPLICANT: Sariasiani, Sina

; APPLICANT: Tang, Xiao-Song

; APPLICANT: Qi, Wei Wei

; TITLE OF INVENTION: Methods for the Production of Tyrosine, Cinnamic Acid and Para-

; FILE REFERENCE: CL1777

; CURRENT APPLICATION NUMBER: US/10/138,970A

; CURRENT FILING DATE: 2002-07-23

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 1

; LENGTH: 894

; TYPE: DNA

; ORGANISM: C. violaceum

US-10-138-970A-1

Query Match 79.0%; Score 15.8; DB 15; Length 894;
Best Local Similarity 89.5%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCTCGACT 19

Db 370 GAGCGCGCACCTCGACT 388

RESULT 26

US-10-156-761-6097

; Sequence 6097, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 243-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 6097

; LENGTH: 1401

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1401)

US-10-156-761-6097

Query Match 79.0%; Score 15.8; DB 15; Length 1401;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCTCGACT 19

Db 565 GCGCGCGCACCTCGACT 583

RESULT 27

US-10-253-971-4

; Sequence 4, Application US/10253971

; Publication No. US20030070192A1

; GENERAL INFORMATION:

; APPLICANT: GEORGES, FAKZY

; APPLICANT: DONG, JIN-ZHUO

; APPLICANT: KELLER, WILF

; APPLICANT: HUSSAIN, ATTA A. K.

; APPLICANT: SELVARAJ, GOPALAN

; APPLICANT: DATLA, RAJU

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING SECONDARY PLANT

; FILE REFERENCE: 73776-159D

; CURRENT APPLICATION NUMBER: US/10/253,971

; CURRENT FILING DATE: 2002-09-25

; PRIOR APPLICATION NUMBER: US'60/072156

; PRIOR FILING DATE: 1998-01-22

; PRIOR APPLICATION NUMBER: US 09/012453

; PRIOR FILING DATE: 1998-01-23

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 1641

; TYPE: DNA

; ORGANISM: Arthrobacter pascens

US-10-253-971-4

Query Match 79.0%; Score 15.8; DB 15; Length 1641;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGCGCACCTCGACT 19

Db 436 GCGCGCGCACCTCGACT 454

RESULT 28

US-10-425-114-23110

; Sequence 23110, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 23110
; LENGTH: 1747
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3594-014-F8_FLI
US-10-425-114-23110

Query Match 79.0%; Score 15.8; DB 13; Length 1747;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGGCACCACTCGACT 19
|||
Db 291 GCGCGGCACCTCGAAT 309

RESULT 29

US-10-425-114-13600
; Sequence 13600, Application US/10425114
; Publication No. US20040034888A1

GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 13600

; LENGTH: 1776

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB143-011-G1_FLI

US-10-425-114-13600

Query Match 79.0%; Score 15.8; DB 13; Length 1776;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGGCACCACTCGACT 19
|||
Db 286 GCGCGGCACCTCGAAT 304

RESULT 30

US-10-425-114-26722
; Sequence 26722, Application US/10425114
; Publication No. US20040034888A1

GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 26722
; LENGTH: 1781
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4572-009-E7_FLI
US-10-425-114-26722

Query Match 79.0%; Score 15.8; DB 13; Length 1781;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGCGGCACCACTCGACT 19
|||
Db 325 GCGCGGCACCTCGAAT 343

Search completed: June 20, 2004, 17:45:46
Job time : 131.122 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 597.376 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22

Sequence: 1 ccgggcaaaacctcgctaac 22

Scoring table: IDENTITY NUC

Gap 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	22	100.0	335	1	MBO298910	AJ298910 Mycobacte
C 2	22	100.0	335	1	MBO298911	AJ298911 Mycobacte
C 3	22	100.0	390	1	NM1298907	AJ298907 Mycobacte
C 4	22	100.0	390	1	NM1298909	AJ298909 Mycobacte
C 5	22	100.0	414	1	MBO298905	AJ298905 Mycobacte
C 6	22	100.0	469	1	MTU298905	AJ298906 Mycobacte
C 7	22	100.0	493	1	MBO298908	AJ298908 Mycobacte
C 8	22	100.0	10894	1	AE007121	AE007121 Mycobacte
C 9	22	100.0	318050	1	BX248344	BX248344 Mycobacte
C 10	22	100.0	348676	15	EX842581	EX842581 Mycobacte
C 11	18.8	85.5	242139	2	AC095431	AC095431 Rattus no
C 12	17.8	80.9	2629	3	AV119652	AV119652 Drosophil
C 13	17.8	80.9	46594	2	AC013897	AC013897 Drosophil
C 14	17.8	80.9	164415	3	AC023718	AC023718 Drosophil
C 15	17.8	80.9	166296	2	AC102929	AC102929 Mus muscu
C 16	17.8	80.9	229648	2	AC117358	AC117358 Rattus no
C 17	17.8	80.9	242397	2	AC097187	AC097187 Rattus no
C 18	17.8	80.9	254540	2	AC119383	AC119383 Rattus no
C 19	17.8	80.9	260335	2	AC115140	AC115140 Rattus no
C 20	17.8	80.9	286789	2	AC133723	AC133723 Rattus no
C 21	17.8	80.9	328128	3	AE003486	AE003486 Drosophil
C 22	17.4	79.1	163448	8	OSJN00127	AL606999 Oryza sat
C 23	17.4	79.1	176859	2	BX510331	BX510331 Danio rer
C 24	17.4	79.1	325612	2	AC095761	AC095761 Rattus no
C 25	17.2	78.2	2171	1	SEU43356	U43356 Salmonella
C 26	17.2	78.2	2171	1	SEU43357	U43357 Salmonella
C 27	17.2	78.2	2467	1	SEU43350	U43350 Salmonella
C 28	17.2	78.2	2467	1	SEU43351	U43351 Salmonella
C 29	17.2	78.2	2468	1	SEU43346	U43346 Salmonella
C 30	17.2	78.2	2468	1	SEU43347	U43347 Salmonella
C 31	17.2	78.2	5925	6	AX346478	AX346478 Sequence
C 32	17.2	78.2	64011	9	AC007642	AC007642 Homo sapi
C 33	17.2	78.2	65005	2	AC091277	AC091277 Mus muscu
C 34	17.2	78.2	77292	8	EX842596	EX842596 Neurospor
C 35	17.2	78.2	93508	10	AL606926	AL606926 Mouse DNA
C 36	17.2	78.2	189761	10	AC131177	AC131177 Mus muscu
C 37	17.2	78.2	192380	2	AC140346	AC140346 Mus muscu
C 38	17.2	78.2	204133	10	AC124504	AC124504 Mus muscu
C 39	17.2	78.2	217281	2	AC142259	AC142259 Mus muscu
C 40	17.2	78.2	217490	10	AC132402	AC132402 Mus muscu
C 41	17.2	78.2	225260	2	AC095002	AC095002 Rattus no
C 42	17.2	78.2	229299	2	AC097991	AC097991 Rattus no
C 43	17.2	78.2	262050	1	AP000985	AP000985 Sulfolobu
C 44	17.2	78.2	316869	2	AC124509	AC124509 Mus muscu
C 45	17.2	78.2	319431	2	AC097106	AC097106 Rattus no
C 46	17	77.3	46530	2	EX293554_5	Continuation (6 of
C 47	17	77.3	123556	9	HS071115	Z98950 Human DNA s
C 48	16.8	76.4	536	6	AX307280	AX307280 Sequence
C 49	16.8	76.4	6054	1	CANRDFGEN	Y09572 Corynebacte
C 50	16.8	76.4	77500	8	AP004166	AP004166 Oryza sat
C 51	16.8	76.4	84113	9	AC073063	AC073063 Homo sapi
C 52	16.8	76.4	156598	5	EX005055	EX005055 Zebrafish
C 53	16.8	76.4	185871	10	AC074311	AC074311 Mus muscu
C 54	16.8	76.4	219339	2	AC112623	AC112623 Rattus no
C 55	16.8	76.4	220280	2	AC145981	AC145981 Pan trogl
C 56	16.8	76.4	233152	2	AC134085	AC134085 Rattus no
C 57	16.8	76.4	237115	2	AC109127	AC109127 Rattus no
C 58	16.8	76.4	242845	2	AC126976	AC126976 Rattus no
C 59	16.8	76.4	246099	2	AC106365	AC106365 Rattus no
C 60	16.8	76.4	247570	2	AC130862	AC130862 Rattus no
C 61	16.8	76.4	259226	2	AC133656	AC133656 Rattus no
C 62	16.8	76.4	260199	2	AC110409	AC110409 Rattus no
C 63	16.8	76.4	279112	2	AC132048	AC132048 Rattus no
C 64	16.8	76.4	288925	2	AC146422	AC146422 Pan trogl
C 65	16.8	76.4	303846	2	AC103292	AC103292 Rattus no

66	16.8	76.4	340900	1	SME591791	AL591791 Sinorhizo	139	16.2	73.6	96482	2	AC136041	AC136041 Rattus no
67	16.4	74.5	771	6	AX677905	AX677905 Sequence	c 140	16.2	73.6	97774	5	AL591399	AL591399 Zebrafish
68	16.4	74.5	2366	8	AK103682	AK103682 Oryza sat	c 141	16.2	73.6	110000	0	AC099174	AC099174 Rattus no
69	16.4	74.5	3000	1	R2TKILA	M62846 Plasmid RK2	c 142	16.2	73.6	110000	3	AE001572	AE001572 Contamination (2 of
70	16.4	74.5	3086	1	K2KILA	M62426 Plasmid RK2	c 143	16.2	73.6	112732	9	AL160007	AL160007 Human DNA
71	16.4	74.5	3502	1	NGAJ2783	AJ002783 Neisseria	c 144	16.2	73.6	113164	2	AC136925	AC136925 Oryza sat
72	16.4	74.5	3738	9	BC005836	BC005836 Homo sapi	c 145	16.2	73.6	119014	2	AC142510	AC142510 Rattus no
73	16.4	74.5	3832	9	HSU12535	UI2535 Human epide	c 146	16.2	73.6	119909	2	AC068935	AC068935 Homo sapi
74	16.4	74.5	3887	9	BC030010	BC030010 Homo sapi	c 147	16.2	73.6	124921	2	AC133132	AC133132 Rattus no
75	16.4	74.5	10029	1	AE015044	AE015044 Shigella	c 148	16.2	73.6	130913	2	AC096869	AC096869 Rattus no
76	16.4	74.5	12106	1	AE001497	AE001497 Helicobac	c 149	16.2	73.6	138416	9	AC105218	AC105218 Homo sapi
77	16.4	74.5	12167	1	AE005184	AE005184 Escherich	c 150	16.2	73.6	144111	8	AP002481	AP002481 Oryza sat
78	16.4	74.5	14376	1	AE002531	AE002531 Neisseria	c 151	16.2	73.6	144094	9	AC087209	AC087209 Homo sapi
79	16.4	74.5	28043	2	AC014190	AC014190 Drosophil	c 152	16.2	73.6	145437	2	AC139985	AC139985 Rattus no
80	16.4	74.5	60099	1	BIACONGEN	L27758 Birmingham	c 153	16.2	73.6	156338	9	AC022887	AC022887 Homo sapi
81	16.4	74.5	64790	8	AP004863	AP004863 Oryza sat	c 154	16.2	73.6	158792	2	AC044874	AC044874 Homo sapi
82	16.4	74.5	75282	2	DMBR37M19	AL133495 Drosophil	c 155	16.2	73.6	165797	2	AC141664	AC141664 Apis mell
83	16.4	74.5	84353	9	AC092753	AC092753 Homo sapi	c 156	16.2	73.6	166015	9	AC113194	AC113194 Homo sapi
84	16.4	74.5	123354	8	AC123518	AC123518 Oryza sat	c 157	16.2	73.6	172937	10	AC121914	AC121914 Mus muscu
85	16.4	74.5	134011	9	AC005886	AC005886 b240916,	c 158	16.2	73.6	177227	2	AC103580	AC103580 Trypanoso
86	16.4	74.5	141534	2	AP003685	AP003685 Oryza sat	c 159	16.2	73.6	181845	2	AC026586	AC026586 Homo sapi
87	16.4	74.5	162219	8	AC145321	AC145321 Oryza sat	c 160	16.2	73.6	182556	9	AC016879	AC016879 Homo sapi
88	16.4	74.5	163953	2	BX000689	BX000689 Danio rer	c 161	16.2	73.6	186107	2	AC023132	AC023132 Homo sapi
89	16.4	74.5	179009	9	AL355543	AL355543 Human DNA	c 162	16.2	73.6	186582	2	AC127620	AC127620 Rattus no
90	16.4	74.5	179892	3	AC023691	AC023691 Drosophil	c 163	16.2	73.6	187515	2	AC125383	AC125383 Rattus no
91	16.4	74.5	187223	2	AC008333	AC008333 Drosophil	c 164	16.2	73.6	190512	2	AC135699	AC135699 Rattus no
92	16.4	74.5	222746	2	EX510345	EX510345 Danio rer	c 165	16.2	73.6	190542	3	AC095015	AC095015 Drosophil
93	16.4	74.5	228802	2	AC023717	AC023717 Drosophil	c 166	16.2	73.6	199412	10	AC122342	AC122342 Mus muscu
94	16.4	74.5	251952	2	AC094693	AC094693 Rattus no	c 167	16.2	73.6	201415	2	AC117103	AC117103 Rattus no
95	16.4	74.5	281530	1	AP002550	AP002550 Escherich	c 168	16.2	73.6	202050	1	AL646065	AL646065 Ralstonia
96	16.4	74.5	290029	1	AE016978	AE016978 Shigella	c 169	16.2	73.6	203223	2	AC116077	AC116077 Rattus no
97	16.4	74.5	300409	1	AE016755	AE016755 Escherich	c 170	16.2	73.6	216562	2	AC130262	AC130262 Rattus no
98	16.4	74.5	308217	3	AE003487	AE003487 Drosophil	c 171	16.2	73.6	221118	2	AC095855	AC095855 Rattus no
99	16.4	74.5	323069	2	AC079737	AC079737 Homo sapi	c 172	16.2	73.6	223438	10	AL662804	AL662804 Mouse DNA
100	16.4	74.5	349980	6	AX044034	AX044034 Sequence	c 173	16.2	73.6	238055	2	AC131139	AC131139 Rattus no
101	16.2	73.6	462	10	RATATM2	M34138 Rat alpha-t	c 174	16.2	73.6	238055	2	AC131139	AC131139 Rattus no
102	16.2	73.6	753	6	BD246045	BD246045 Developme	c 175	16.2	73.6	239021	2	AC102973	AC102973 Rattus no
103	16.2	73.6	794	1	AX134487	AX134487 Providenc	c 176	16.2	73.6	240652	2	AC111587	AC111587 Rattus no
104	16.2	73.6	812	6	AX109605	AX109605 Sequence	c 177	16.2	73.6	240890	2	AC123128	AC123128 Rattus no
105	16.2	73.6	819	6	AX109606	AX109606 Sequence	c 178	16.2	73.6	245023	2	AC129467	AC129467 Rattus no
106	16.2	73.6	830	6	AX109604	AX109604 Sequence	c 179	16.2	73.6	245524	2	AC095456	AC095456 Rattus no
107	16.2	73.6	1050	10	RATMBRIA	M34135 Rat brain a	c 180	16.2	73.6	245654	2	AC116205	AC116205 Rattus no
108	16.2	73.6	1241	8	AX058156	AX058156 Arabidops	c 181	16.2	73.6	245907	2	AC095162	AC095162 Rattus no
109	16.2	73.6	1362	8	AX066816	AX066816 Oryza sat	c 182	16.2	73.6	251073	2	AC108534	AC108534 Rattus no
110	16.2	73.6	1366	3	AC006700	AC006700 Caenorhab	c 183	16.2	73.6	251244	2	AC103149	AC103149 Rattus no
111	16.2	73.6	1911	6	BD260923	BD260923 Novel met	c 184	16.2	73.6	252192	2	AC095521	AC095521 Rattus no
112	16.2	73.6	1911	6	AX036306	AX036306 Sequence	c 185	16.2	73.6	253021	2	AC103467	AC103467 Rattus no
113	16.2	73.6	1911	6	AX038916	AX038916 Sequence	c 186	16.2	73.6	254244	2	AC094541	AC094541 Rattus no
114	16.2	73.6	2488	10	RATMBRIA	M34136 Rat brain a	c 187	16.2	73.6	256703	2	AC114076	AC114076 Rattus no
115	16.2	73.6	3773	3	DNAC001656	AC001656 Drosophil	c 188	16.2	73.6	257071	1	AP002558	AP002558 Escherich
116	16.2	73.6	6022	6	AX251151	AX251151 Sequence	c 189	16.2	73.6	261089	2	AC094807	AC094807 Rattus no
117	16.2	73.6	6022	6	AX277916	AX277916 Sequence	c 190	16.2	73.6	261143	2	AC097135	AC097135 Rattus no
118	16.2	73.6	6022	6	AX323597	AX323597 Sequence	c 191	16.2	73.6	261434	2	AC130556	AC130556 Rattus no
119	16.2	73.6	6022	6	AX345974	AX345974 Sequence	c 192	16.2	73.6	270712	2	AC137360	AC137360 Rattus no
120	16.2	73.6	9888	6	AX346143	AX346143 Sequence	c 193	16.2	73.6	271725	2	AC116298	AC116298 Rattus no
121	16.2	73.6	10358	1	AE005403	AE005403 Escherich	c 194	16.2	73.6	290628	1	AE016983	AE016983 Shigella
122	16.2	73.6	10378	1	AE000275	AE000275 Escherich	c 195	16.2	73.6	299864	2	AC006702	AC006702 Caenorhab
123	16.2	73.6	10764	1	AE015166	AE015166 Shigella	c 196	16.2	73.6	300475	1	AE016761	AE016761 Escherich
124	16.2	73.6	10814	1	AE014472	AE014472 Brucella	c 197	16.2	73.6	303446	1	AE016941	AE016941 Bacteroid
125	16.2	73.6	11987	1	AE009465	AE009465 Brucella	c 198	16.2	73.6	304250	1	EX294152	EX294152 Pirellula
126	16.2	73.6	16668	6	BD246036	BD246036 Developme	c 199	16.2	73.6	308146	3	AE003673	AE003673 Drosophil
127	16.2	73.6	16784	7	AF513032	AF513032 Staphyloc	c 200	16.2	73.6	313735	2	AC094115	AC094115 Rattus no
128	16.2	73.6	16897	1	D90825	D90825 E.coli geno	c 201	16	72.7	998	11	CNS06F74	AL396038 T7 end of
129	16.2	73.6	18227	7	AF513033	AF513033 Staphyloc	c 202	16	72.7	9874	1	AE011350	AE011350 Leptospir
130	16.2	73.6	32179	3	AF047663	AF047663 Caenorhab	c 203	16	72.7	124215	2	AC143118	AC143118 Macaca mu
131	16.2	73.6	36579	2	AC134070	AC134070 Rattus no	c 204	16	72.7	151806	8	AP002093	AP002093 Oryza sat
132	16.2	73.6	56043	2	AC012761	AC012761 Drosophil	c 205	16	72.7	300050	1	AP004599	AP004599 Oceanobac
133	16.2	73.6	58464	2	AC110049	AC110049 Homo sapi	c 206	15.8	71.8	20	6	AX297024	AX297024 Sequence
134	16.2	73.6	74045	9	EX284671	EX284671 Human DNA	c 207	15.8	71.8	24	6	AX292391	AX292391 Sequence
135	16.2	73.6	80866	9	AC012649	AC012649 Drosophil	c 208	15.8	71.8	775	9	HS334076	HS334076 Homo sapi
136	16.2	73.6	81662	8	AB008265	AB008265 Arabidops	c 209	15.8	71.8	954	6	AR096464	AR096464 Sequence
137	16.2	73.6	87102	2	AL606983	AL606983 Danio rer	c 210	15.8	71.8	1101	6	AX459691	AX459691 Sequence
138	16.2	73.6	95968	3	CEY47D3B	AL031635 Caenorhab	c 211	15.8	71.8	1280	1	RHMN1FHW	K01620 Rhizobium j


```

212 15.8 71.8 1281 6 E00713 Genomic DNA
213 15.8 71.8 1347 6 AR096466
214 15.8 71.8 1587 5 US9421 Gallus gall
215 15.8 71.8 1587 5 AR096465
216 15.8 71.8 2233 6 AR096463
217 15.8 71.8 2236 1 STYCIITCA
218 15.8 71.8 3188 3 BT001752
219 15.8 71.8 3795 3 BT003565
220 15.8 71.8 5027 6 AR096462
221 15.8 71.8 8391 1 AY392415
222 15.8 71.8 11916 3 DME249115
223 15.8 71.8 13124 1 AE009105
224 15.8 71.8 13147 1 AE008070
225 15.8 71.8 13639 1 BS293937
226 15.8 71.8 19165 2 AC134163
227 15.8 71.8 22507 6 AX702487
228 15.8 71.8 27923 2 AC095226
229 15.8 71.8 35737 9 AC005263
230 15.8 71.8 39143 9 AC004490
231 15.8 71.8 49656 2 AC104957
232 15.8 71.8 53038 2 AC111637
233 15.8 71.8 60894 2 AC090930
234 15.8 71.8 62749 2 AC019842
235 15.8 71.8 74759 2 AC100661
236 15.8 71.8 77283 2 AC020039
237 15.8 71.8 78290 2 AC108631
238 15.8 71.8 83412 2 AC133611
239 15.8 71.8 89246 2 AC068226
240 15.8 71.8 90502 9 AC108676
241 15.8 71.8 93829 3 AC007112
242 15.8 71.8 95419 2 AC133704
243 15.8 71.8 98955 9 AC008886
244 15.8 71.8 99204 2 AC135098
245 15.8 71.8 101033 2 AC136177
246 15.8 71.8 110000 2 AC092921
247 15.8 71.8 110000 2 AC095226
248 15.8 71.8 110000 2 AC097907
249 15.8 71.8 110000 2 AC107271
250 15.8 71.8 110000 2 AC107271
251 15.8 71.8 110000 2 AC113863
252 15.8 71.8 110000 2 AC115269
253 15.8 71.8 110000 2 AC118411
254 15.8 71.8 110000 2 AC125991
255 15.8 71.8 110000 2 AC125991
256 15.8 71.8 110000 2 AC127799
257 15.8 71.8 110000 2 AC129112
258 15.8 71.8 110000 2 AC132558
259 15.8 71.8 110000 2 LMFLCHR15_03
260 15.8 71.8 114134 9 AF236876
261 15.8 71.8 115525 9 AC117491
262 15.8 71.8 115597 2 AC136166
263 15.8 71.8 118994 9 AL139823
264 15.8 71.8 121414 2 AC017234
265 15.8 71.8 128561 2 AC068729
266 15.8 71.8 129957 9 HS415G2
267 15.8 71.8 137061 2 AC136046
268 15.8 71.8 138719 2 AC126534
269 15.8 71.8 143897 9 AC011647
270 15.8 71.8 144552 8 CNS08CBX
271 15.8 71.8 151211 2 AC117337
272 15.8 71.8 151321 2 AC136264
273 15.8 71.8 152664 2 AC128855
274 15.8 71.8 155592 2 AC140736
275 15.8 71.8 157419 2 AC025389
276 15.8 71.8 160287 2 AC124150
277 15.8 71.8 161177 2 AC125850
278 15.8 71.8 163308 2 AC137174
279 15.8 71.8 165051 9 AC026950
280 15.8 71.8 165190 3 AC093096
281 15.8 71.8 165304 2 AC135746
282 15.8 71.8 165765 2 AC117979
283 15.8 71.8 166504 9 AC008275
284 15.8 71.8 167243 2 AC125956

c 285 15.8 71.8 167841 2 AC125767
c 286 15.8 71.8 169226 6 AC073309
c 287 15.8 71.8 169835 9 AC020896
c 288 15.8 71.8 170040 2 AC141975
c 289 15.8 71.8 173165 3 AC099005
c 290 15.8 71.8 173783 2 AC117135
c 291 15.8 71.8 173965 9 AL159977
c 292 15.8 71.8 175091 2 AC141991
c 293 15.8 71.8 176024 2 CNS09SC3
c 294 15.8 71.8 177109 2 AC118161
c 295 15.8 71.8 177278 3 AC007726
c 296 15.8 71.8 177787 9 AC133961
c 297 15.8 71.8 178778 2 AC105810
c 298 15.8 71.8 179115 2 AC097862
c 299 15.8 71.8 180000 1 AF222012
c 300 15.8 71.8 180290 2 AC123434

RESULT 1
LOCUS MBO298910/c 335 bp DNA linear BCT 23-OCT-2001
DEFINITION Mycobacterium bovis VNTR, strain 1503/96.
ACCESSION AJ298910
VERSION AJ298910.1 GI:16416365
KEYWORDS repetitive element; tandem repeat; VNTR.
SOURCE Mycobacterium bovis
ORGANISM Mycobacterium bovis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
1 Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
Fluorescent VNTR (fVNTR) for improved high throughput molecular
typing of Mycobacterium bovis
Unpublished
2 (bases 1 to 335)
Llewellyn-Hughes, S.
Direct Submission
Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
Surrey, KT15 3NB, UNITED KINGDOM
FEATURES
source
1..335
/organism="Mycobacterium bovis"
/mol_type="genomic DNA"
/strain="1503/96"
/db_xref="taxon:1765"
repeat_region
73..127
/rpt_family="VNTR F2.1"
/rpt_type="TANDEM"
repeat_region
128..206
/rpt_family="VNTR F2.1"
/rpt_type="TANDEM"
repeat_region
207..292
/rpt_family="VNTR F2.1"
/rpt_type="TANDEM"
ORIGIN
Query Match 100.0%; Score 22; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 8.1; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;
QY 1 CCGGGCAAAACCTCGCCTAAC 22
|||||
Db 317 CCGGGCAAAACCTCGCCTAAC 296
|||||

RESULT 2
MBO298911/c 335 bp DNA linear BCT 23-OCT-2001
LOCUS
```

```

DEFINITION   Mycobacterium bovis VNTR, strain 4705.
ACCESSION    AJ298911
VERSION      AJ298911.1 GI:16416366
KEYWORDS     repetitive element; tandem repeat; VNTR.
SOURCE       Mycobacterium bovis
ORGANISM     Mycobacterium bovis
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 335)
AUTHORS      Llewellyn-Hughes, S.
TITLE        Direct Submission
JOURNAL      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
              Location/Qualifiers
              1..335
              /organism="Mycobacterium bovis"
              /mol_type="genomic DNA"
              /strain="4705"
              /db_xref="taxon:1765"
              73..127
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
              128..206
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
              207..292
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM

repeat_region
              73..127
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM

repeat_region
              128..206
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM

repeat_region
              207..292
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM

ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 CCGGCAAAACCTCGGCTAAC 22
    |||||
Db 317 CCGGCAAAACCTCGGCTAAC 296
    |||||

RESULT 3
MMI298907/c
LOCUS           MMI298907 390 bp DNA linear BCT 23-OCT-2001
DEFINITION     Mycobacterium microti VNTR, strain 4777/97.
ACCESSION      AJ298907
VERSION        AJ298907.1 GI:16416362
KEYWORDS       repetitive element; tandem repeat; VNTR.
SOURCE         Mycobacterium microti
ORGANISM       Mycobacterium microti
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 390)
AUTHORS      Llewellyn-Hughes, S.
TITLE        Direct Submission
JOURNAL      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
              Location/Qualifiers
              1..390
              /organism="Mycobacterium microti"
              /mol_type="genomic DNA"
              /strain="4700/97"
              /db_xref="taxon:1806"
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM

repeat_region
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM

repeat_region
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM

repeat_region
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM

ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 CCGGCAAAACCTCGGCTAAC 22
    |||||
Db 372 CCGGCAAAACCTCGGCTAAC 351
    |||||

RESULT 4
MMI298909/c
LOCUS           MMI298909 390 bp DNA linear BCT 23-OCT-2001
DEFINITION     Mycobacterium microti VNTR, strain 4700/97.
ACCESSION      AJ298909
VERSION        AJ298909.1 GI:16416364
KEYWORDS       repetitive element; tandem repeat; VNTR.
SOURCE         Mycobacterium microti
ORGANISM       Mycobacterium microti
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE    1
AUTHORS      Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE        Fluorescent VNTR (fVNTR) for improved high throughput molecular
              typing of Mycobacterium bovis
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 390)
AUTHORS      Llewellyn-Hughes, S.
TITLE        Direct Submission
JOURNAL      Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
              Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
              Surrey, KT15 3NB, UNITED KINGDOM
              Location/Qualifiers
              1..390
              /organism="Mycobacterium microti"
              /mol_type="genomic DNA"
              /strain="4700/97"
              /db_xref="taxon:1806"
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM

repeat_region
              73..127
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM

repeat_region
              128..182
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM

repeat_region
              183..261
              /rpt_family="VNTR F2.2"
              /rpt_type=TANDEM

ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 390;
Best Local Similarity 100.0%; Pred. No. 8.1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 CCGGCAAAACCTCGGCTAAC 22
    |||||
Db 372 CCGGCAAAACCTCGGCTAAC 351
    |||||

RESULT 5
MMI298905/c

```

```

LOCUS      MBO298905          414 bp    DNA          linear      BCT 23-OCT-2001
DEFINITION Mycobacterium bovis VNTR, strain AN 5.
ACCESSION  AJ298905
VERSION    AJ298905.1 GI:16416360
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium bovis
ORGANISM   Mycobacterium bovis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE  1
AUTHORS    Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE      Fluorescent VNTR (fVNTR) for improved high throughput molecular
            typing of Mycobacterium bovis
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 414)
AUTHORS    Llewellyn-Hughes, S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
            Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
            Surrey, KT15 3NB, UNITED KINGDOM
FEATURES   Location/Qualifiers
            source
              1..414
              /organism="Mycobacterium bovis"
              /mol_type="genomic DNA"
              /strain="AN 5"
              /db_xref="taxon:1765"
            repeat_region
              73..127
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
            repeat_region
              128..206
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM
            repeat_region
              207..292
              /rpt_family="VNTR F2.1"
              /rpt_type=TANDEM

ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 414;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  CCGGGCAAAACCTCGGCTTAC 22
        |||||
        396 CCGGGCAAAACCTCGGCTTAC 375

DB

RESULT 6
LOCUS      MTU298906          469 bp    DNA          linear      BCT 23-OCT-2001
DEFINITION Mycobacterium tuberculosis VNTR, strain H37Rv.
ACCESSION  AJ298906
VERSION    AJ298906.1 GI:16416361
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium tuberculosis
ORGANISM   Mycobacterium tuberculosis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE  1
AUTHORS    Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE      Fluorescent VNTR (fVNTR) for improved high throughput molecular
            typing of Mycobacterium bovis
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 469)
AUTHORS    Llewellyn-Hughes, S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
            Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
            Surrey, KT15 3NB, UNITED KINGDOM
FEATURES   Location/Qualifiers
            source
              1..469
              /organism="Mycobacterium tuberculosis"

```

```

/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:1773"
73..127
/rpt_family="VNTR F2.1"
/rpt_type=TANDEM
128..206
/rpt_family="VNTR F2.1"
/rpt_type=TANDEM
207..292
/rpt_family="VNTR F2.1"
/rpt_type=TANDEM

ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 469;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  CCGGGCAAAACCTCGGCTTAC 22
        |||||
        451 CCGGGCAAAACCTCGGCTTAC 430

DB

RESULT 7
LOCUS      MBO298908/c          493 bp    DNA          linear      BCT 23-OCT-2001
DEFINITION Mycobacterium bovis VNTR, strain 2259-93.
ACCESSION  AJ298908
VERSION    AJ298908.1 GI:16416363
KEYWORDS   repetitive element; tandem repeat; VNTR.
SOURCE     Mycobacterium bovis
ORGANISM   Mycobacterium bovis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE  1
AUTHORS    Llewellyn-Hughes, S., Llewellyn-Hughes, J. and Hewinson, R.G.
TITLE      Fluorescent VNTR (fVNTR) for improved high throughput molecular
            typing of Mycobacterium bovis
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 493)
AUTHORS    Llewellyn-Hughes, S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-2000) Llewellyn-Hughes S., Dept. of Bacterial
            Diseases, Veterinary Laboratories Agency, New Haw, Addlestone,
            Surrey, KT15 3NB, UNITED KINGDOM
FEATURES   Location/Qualifiers
            source
              1..493
              /organism="Mycobacterium bovis"
              /mol_type="genomic DNA"
              /strain="2259-93"
              /db_xref="taxon:1765"
            repeat_region
              73..127
              /rpt_family="VNTR F4.1"
              /rpt_type=TANDEM
            repeat_region
              128..206
              /rpt_family="VNTR F4.1"
              /rpt_type=TANDEM
            repeat_region
              207..285
              /rpt_family="VNTR F4.1"
              /rpt_type=TANDEM

ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 493;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  CCGGGCAAAACCTCGGCTTAC 22
        |||||
        475 CCGGGCAAAACCTCGGCTTAC 454

DB

RESULT 8

```

AE007121/c
LOCUS 10894 bp DNA linear BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 207 of 280 of the complete genome.
ACCESSION AE007121 AB000516
VERSION AE007121.1 GI:13882769
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacterii; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 10894)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Winn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
Unpublished
2 (bases 1 to 10894)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Winn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA
FEATURES
source
1..10894
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/notes="clinical strain"
complement(247..1515)
/gene="WT2989"
complement(247..1515)
/notes="WT2989"
/notes="similar to PID:1237016 SP:P51835 GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="signal recognition particle FtsY"
/protein_id="AAK47316.1"
/db_xref="GI:13882770"
/translation="MWELMTATAVIAALVIAALTLGLVLRRLRISLSPRPGVW
DRSGYTAGSGITTSQPTTQPAERIDTSGLPVAGDDATVPDAPKRTIADVHLPRFE
PEQAPVEPADALAPGRLRLRGLRSLQNALGRLGLGGLDSDSDQVYED
TLIVADGPATASVQLSRSLASGNVREADARAVLRVLINELOPQMDRSIRALP
HAGPSVLLVNGVGTGKTTVTKGLRVLDVAGRRVVGAAATFRAAAALQLOQTWAAR
VGAAVVRPEGADPASVAFADVDKGIAGADVWLDITAGRLHTKVGIMDELKVRV
TRRASVDELVLVDATIGQGLAQARFAEVFVDSGAVLTKLDGTAKGGIVFRVQOEL
GVPVKLVGLGPGDDLPAPFAFVALLG"
complement(1565..5182)
/gene="WT2990"
complement(1565..5182)
/notes="similar to PID:1237015 SP:P51834 GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chromosome segregation SMC protein, putative"
/protein_id="AAK47317.1"
/db_xref="GI:13882771"
/translation="MYLKSUTLKGFKSAFPTTLRPFPGITAVVNGSGKSNVVDAL
AWNGEQATLGRGKNGVIFAGTSRPLGRAEVTVSDNSDALPIETVEVSTR
RMFRDGASVEYINGSCRLMDVGLSDSGIGRHEMVIVQGGKLEILQSRPDRRAF
IEEAAGVLRKRLKRLDWTMAANLRLTDLTTLRQLKPLGRQAEAAQRAAI
QADLRDLRLAADDLVSRAREAVFQAEANMEREDAEAAELAVASBELAHASAV
AEJSTRASIQHTWFGISALAEVDAVRIASERAHILDIETPVANSDTPRKPELEA

EAQVAVAEQQLLAELDARASILDARAEADRAAEADRAHAAVVEEADREGL
ARLQGVETMARVESIDSVARLSERIEDAARAOQTRAEFTVQGRIGELDQGEVG
LDHHERTVAAURLVDSVAERQVARSURADALAVGLQSRKQGAALAH
NRSAGLFGSIAQLVKVRYEAAALGPAADALVDGLTAAGSAVSKAODGGR
AVLWSDPAPQASGEMLPSCAQWALDVEGPPQLVGAMISLGVAVVNDLTE
AMGLVIRPELRVAVTDGLVGAVGSGSDKSLTLEVTSIDKARSELAAEALAA
QNLAAAGALTQESAPQDAEALALNESDTAISAMVBOICRIGCEARAEENRL
LOOTFEAVRTQTLDDVTLOLTKAGETORVOAOPIDRCAISAADARGEVE
ARLAVTAERANVRGRADSLRRRAAAREARVRAQQAARLHAHAAVAACGR
LIAGRUHRAVDGASQLRDASAAQROORLAAMAVRVNTLSARVGLTDSUHRDLA
NAQAURIEQLEQVLEQFMAADLITTEYHPHALPTELEMAFEQAREGEVIA
PAPMPDRVTOERRAKRAELAEIGRNVLALEEFALAEERFNLQLEDEKAAR
DLIGVADVDAIRILOVFNDAFVEREFGVFTALPGEGRLRLTEPDMLTGIEV
EARPPQKITRLSLGSGKALTAVAMLVAFRAPSFPYIMDEAALDDVNLRLRL
SLPEQLREQSQIIITHQKPTMEVADALYGTVMQNDGITAVISQRMQGOVDLVNS
S"
complement(5194..5475)
/gene="WT2991"
complement(5194..5475)
/notes="identified by match to protein family HMW"
/codon_start=1
/transl_table=11
/product="acylphosphatase"
/protein_id="AAK47318.1"
/db_xref="GI:13882772"
/translation="MSAPDVRILTAVHWGVQGVGRWTRCRALGLTGYAANHADG
RVLVVAQGPRAACKLLOLQDQTTTPGRVAKVVDWSQSTQITGFSER"
complement(5462..5875)
/gene="WT2992"
complement(5462..5875)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47319.1"
/db_xref="GI:13882773"
/translation="MTQLWERTGTRRYIGRSTRGAQVLVGSDEVGVFTPEGLKTA
LAACSGMASDOPLARLDGDDYQAVVKSADRDQERYPLIEETMELDLSGLTDEKE
RLLVINRAVELACTVGRTLKSGTTNLEVVVDGA"
5867..6013
/gene="WT2993"
5867..6013
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK47320.1"
/db_xref="GI:13882774"
/translation="WRHGFPLPGCLVCVIDGSAPARESIALTSMRQHPYFRGRGW
HRGQ"
complement(5977..6846)
/gene="WT2994"
complement(5977..6846)
/notes="similar to GP.3150225; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="formamidopyrimidine-DNA glycosylase"
/protein_id="AAK47321.1"
/db_xref="GI:13882775"
/translation="XPELPEVVRGLOAHVTGRTITEVVRHPRVHRHAGPADL
TARLNGARINQDRGKILMLTNTAGVHRFTDIALVHLGMSQMLLGAVPCAAHVR
ISALLDDTSLFADQRTFGGLLADLVTVDSGVVPPVFAHAEPLDPRFDCDAVVR
VLRKSLKQLLDQVYSGIGNIYADLRVAKVAVPVAHVAATLRCRRLGVAJHAA
DMREAAKAGGTSTFSLVYVNGSGYFERSLDAYGEGENCRGCAVIRERPMNRS
SFYPCRCQPRRK"
complement(7205..7927)
/gene="WT2995"
complement(7205..7927)

```
/gene="Mt2995"
/note="similar to GP:5114183; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribonuclease III"
/protein_id="AAK47322.1"
/db_xref="GI:13882776"
/translation="MIRSQPLDGLGVLDPELLSLATHSYAYENGGLPTNERLE
FLGDVGLTITDALPHRPHDSRSEGLAKRASVNTQALADVARRLCAGGLGVHVL
GREGANTGGKSSILLADGESSLLGAIYIQHGMKAREVILRLFLGLDAAPTLGAGL
DKWTSLOELTAARGLGAPSYLVITGTPDHDKEFTAVVYVVDSEYSGVGRSKAEAQK
AAAAANKALEVLDNAPGKTS"
/complement(7324..8511)
/gene="Mt2996"
/complement(7324..8511)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK47323.1"
/db_xref="GI:13882777"
/translation="MARQHGPTAQRHVASPMVTVDIARLGRRCQAMFELHDTVHSARI
GLLIIADQALLDLRLVESVSEGLVTGTAAPTVCARCLSPVRGRVQVAUTEL
FAYPDATDTEDEDEGRVVDTEIDLEQPIIDAVGLELFPSPVCRPDGFLCPQCGV
PLASEGHRHEQIDPRWAKLVEMLGPESDILRGER"
/complement(8598..9341)
/gene="Mt2997"
/complement(8598..9341)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK47324.1"
/db_xref="GI:13882778"
/translation="MAYRVFEALDELISAIVEARGVMTAGCVPRGVLELIDIK
DAIPGELDQDVLDLQAKTHADSMVSSATEAESILNHARTEDRLSDAK
AQADRMVSEARQSERVADAREAIRATAKREYASVSRAQACDRLIENGNTSY
EKAVOEGIKQEOBLVNEVVAANAESTRLVDTAAEADRLSGCEDIYVDNKLAEFE
EFLNGTSLRSVGRGRHQLRTAAGTHDIYVR"
/gene="Mt2998"
/gene="Mt2998"
/sequence_similarity; putative"
/sequence_similarity; putative"
/transl_table=11
/product="thioesterase"
/protein_id="AAK47325.1"
/db_xref="GI:13882779"
/translation="WMSLARHGPRYGGVNGHSDSSGDAKQAAPTLIYPHAGGTAK
DYAFSREFSDVRIAVQPGQHDRSGLPLESIPTLADEIFAMMKPSRIDDPVAF
Query Match 100.0%; Score 22; DB 1; Length 10894;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCGGCCTAAC 22
Db 7225 CCGGCAAAACCTCGGCCTAAC 7204
RESULT 9
BX248344/c
LOCUS BX248344
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
11/14.
ACCESSION BX248344
VERSION BX248344.1
KEYWORDS complete genome.
```

```
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
Gene
CDS
gene
CDS
```

```
Mycobacterium bovis subsp. bovis AF2122/97
Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
Garnier, T., Biglieri, K., Camus, J.-C., Medina, N., Mansoor, H.,
Pryor, M., Dutoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,
Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L.,
Wheeler, P., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V. and
Hewinson, G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 ( Microbiology )
2 (bases 1 to 318050)
Garnier, T.
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers
1. .318050
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
/complement(99..875)
/locus_tag="Mb2882c"
/complement(99..875)
/locus_tag="Mb2882c"
/EC_number="1.-.-.-"
/note="Mb2882c, -, len: 258 aa. Equivalent to Rv2857c,
len: 258 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 258 aa overlap). Probable short-chain
dehydrogenase/reductase (EC 1.-.-.-), highly similar to
various dehydrogenases e.g. O88068|SCI35.33c PROBABLE
DEHYDROGENASE (SDR FAMILY) from Streptomyces coelicolor
(260 aa), FASTA scores: opt: 1208, E(): 2e-68, (72.35%
identity in 253 aa overlap); Q91376|PAL649 from
Pseudomonas aeruginosa PROBABLE SHORT-CHAIN DEHYDROGENASE
(253 aa), FASTA scores: opt: 569, E(): 2.1e-28, (39.2%
identity in 255 aa overlap); Q9EX74|MLHA SDR-LIKE ENZYME
from Rhodococcus erythropolis (246 aa), FASTA scores: opt:
567, E(): 2.8e-28, (41.15% identity in 248 aa overlap);
etc. Also similar to many Mycobacterium tuberculosis
dehydrogenases e.g. F6B3|RV2002|MT2058|MTCY39.16c
PUTATIVE OXIDOREDUCTASE (260 aa), FASTA score: (38.3%
identity in 248 aa overlap). BELONGS TO THE SHORT-CHAIN
DEHYDROGENASES/REDUCTASES (SDR) FAMILY."
/codon_start=1
/transl_table=11
/product="PROBABLE SHORT-CHAIN TYPE
DEHYDROGENASE/REDUCTASE"
/protein_id="CAD96569.1"
/db_xref="GI:31619629"
/translation="WMDLSQRLAGRVAVITGGSGIGLAAGRRMRAEGATIVVGDVY
EAGGAADDELGLFVPTDVCDBAVNGLFDAETGRIDIAFNAGISPPDNLITEN
TELAQVQVDVNLKSVYCCRAALRHVYLVAGKSGIVNTAFVAVNGSATSIQSYTAS
KGVVLASRELGVQFARQIRNVLCPGPVTPLLQELFAKNPERAARRMVHVLGRF
APEDTAATAVAFASDASDASFTTASFLVDGGISSAVYTP"
/complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"
/complement(872..2239)
/gene="aldC"
/locus_tag="Mb2883c"
```


XX	Mycobacterium tuberculosis H37Rv	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
OS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium;	FT	complement(740..775)
OC	Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
XX		FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
XX		FT	complement(811..846)
RN	MEDLINE; 98295987.	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
RX	PUBMED; 9634230.	FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	FT	complement(884..919)
RA	Gordon S.V., Eigmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
RA	Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,	FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RA	Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,	FT	complement(958..993)
RA	Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
RA	Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,	FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RA	Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;	FT	complement(1031..1066)
RA	"Deciphering the biology of Mycobacterium tuberculosis from the complete	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
RT	genome sequence";	FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RT	Nature 393:537-544(1998).	FT	complement(1104..1139)
XX		FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
XX		FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RN	PUBMED; 12368430.	FT	complement(1176..1211)
RX	Camus J.C., Pryor M.J., Medigue C., Cole S.T.;	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
RA	"Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";	FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RT	Microbiology 148:2967-2973(2002).	FT	complement(1248..1283)
XX		FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
XX		FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RN		FT	complement(1322..1357)
RX	1-348676	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
RA	Parkhill J.;	FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RA	Submitted (11-JUN-1998) to the EMBL/GenBank/DBJ databases.	FT	complement(1396..1431)
RL	Submitted on behalf of the Mycobacterium tuberculosis sequencing and	FT	/note="36 bp direct repeat, 36 out of 36 bp identical to
RL	mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,	FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
RL	Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut	FT	complement(1467..1486)
RL	Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:	FT	/note="20 bp partial direct repeat, CCCCAGAGGGGACGGAAC,
RL	parkhill@sanger.ac.uk	FT	of sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"
XX		FT	complement(1486..2860)
XX		FT	/note="IS6110-11, len: 1375 bp. Insertion sequence IS6110."
CC	Notes:	FT	/insertion seq="IS6110-11"
CC	Details of M. tuberculosis sequencing at the Sanger Centre	FT	complement(1529..2467)
CC	are available on the World Wide Web.	FT	/note="RV2814c, (MTCY16B7.29), len: 312 aa. Probable
CC	(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/)	FT	transposase, highly similar to others e.g.
XX	Key	FT	P37137 RV0796 MTV042.06 PUTATIVE TRANSPOSASE FOR INSERTION
XX	Location/Qualifiers	FT	SEQUENCE ELEMENT IS986/IS6110 from Mycobacterium
PH	source	FT	tuberculosis (328 aa), FASTA scores: opt: 2103, E():
FT	1..348676	FT	6.1e-132, (100.0% identity in 312 aa overlap); etc. Start
FT	/db_xref="taxon:8332"	FT	unlikely."
FT	/mol_type="genomic DNA"	FT	/transl_table=11
FT	/organism="Mycobacterium tuberculosis H37Rv"	FT	/function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION
FT	/strain="H37Rv"	FT	ELEMENT IS6110."
FT	complement(148..183)	FT	/locus tag="RV2814c"
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to	FT	/product="PROBABLE TRANSPOSASE"
FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"	FT	/protein_id="CAB03675.1"
FT	complement(148..4539)	FT	/translation="LITRIADHOGHREPGDGLRWGVSICTQLTELGVPIAPSTYDHD
FT	/note="4392 bp direct repeat region"	FT	INREPSRELRDGLKEHLSRSHVHAANYGVYARKVNLINREGISVARTVERLMTKL
FT	complement(222..257)	FT	LSGTRGKARRITADIPARPADLVQRRFGPPAPNRLWADLTIVSTWAGYAVFT
FT	/note="36 bp direct repeat, 35 out of 36 bp identical to	FT	DAYARRILGWVASTMATSMDLAEQAIWTQEQGLVDKDVHHTDRGSQYTSIRFS
FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"	FT	ERLADAGIOPSVGASTGSSYDINALAETINGLYKTELKPGKWRSDVELATARWVDF
FT	complement(295..333)	FT	NHRLVQVCGDVPVPELEAAVYVQORPAAG"
FT	/note="36 bp direct repeat, 35 out of 36 bp identical to	FT	complement(2464..2790)
FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"	FT	/note="RV2815c, (MTCY16B7.28), len: 108 aa. Probable
FT	complement(374..409)	FT	transposase, identical from aa 51 with P19772 Y12_MYCTU
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to	FT	PUTATIVE TRANSPOSASE (INSERTION ELEMENT IS986) from
FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"	FT	Mycobacterium tuberculosis (59 aa), FASTA scores: opt: 365,
FT	complement(447..482)	FT	E(): 1.1e-19, (96.6% identity in 59 aa overlap); and other
FT	/note="36 bp direct repeat, 35 out of 36 bp identical to	FT	transposases."
FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"	FT	/transl_table=11
FT	complement(519..554)	FT	/function="REQUIRED FOR THE TRANSPOSITION OF THE INSERTION
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to	FT	ELEMENT IS6110."
FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"	FT	/locus tag="RV2815c"
FT	complement(590..625)	FT	/product="PROBABLE TRANSPOSASE"
FT	/note="36 bp direct repeat, 36 out of 36 bp identical to	FT	/protein_id="CAB03633.1"
FT	sequence GTCGTCAGACCCCAAAACCCCGAGAGGGGACGGAAC"	FT	/translation="MSGSSRRYPPELRERAVRMVAEIRGQHDSEWAAISEVARLIGVG
FT	complement(1664..1699)	FT	CAETVKRWVQAVDAGAREGTTTESAEKRLRDNALRRANILKTASAFFAELD

```

FT repeat_unit complement(2845..2860)
FT /note="16 bp partial direct repeat, GTGCTGACGCCAAA, of
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(2901..2936)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(2975..3011)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3049..3084)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3121..3156)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3193..3228)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3266..3301)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3338..3373)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3399..3434)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3476..3511)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3548..3583)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3624..3659)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3701..3736)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"
FT complement(3774..3809)
FT /note="36 bp direct repeat, 36 out of 36 bp identical to
FT sequence GTGCTGACGCCAAAACCCCGAGAGGGGACGGAAC"

Query Match 100.0%; Score 22; DB 15; Length 348676;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCCTAAC 22
Db 120812 CCGGCAAAACCTCGCCTAAC 120791

RESULT 11
AC095431/c LOCUS
DEFINITION Rattus norvegicus clone CH230-7J16, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC095431
VERSION AC095431.11 GI:30467805
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 242139)
AUTHORS Murny D., Marie, Mettler, M., Lee, A., Abranzen, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, N., Cree, A., D'Souza, L.,

```

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Geisreger, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, I., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensen, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Pastermak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plummer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puzo, M., Quito, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.B.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shwartsbeyn, A., Sisson, I., Sitter, C.D., Smajda, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstein, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 242139)
 Worley, K.C.
 Direct Submission
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 242139)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 9, 2003 this sequence version replaced gi:24941182.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

cdna@fruitfly.berkeley.edu.
Location/Qualifiers
1. .2629
/organism="Drosophila melanogaster"

```
1. .2629
/db_xref="taxon:7227"
```

235. 2160
/gene="spbk1"

```
/note="Longest ORF"  
/codon start=1
```

```

/db_xref="GI:21429174"
/db_xref="FLYBASE:FBgn0030300"
/translation="VTANTGSGENLMGNGKTHPEPTSTDEVSAGTPTLSEIFF
VNGRRKQSIKIQVKLEPGVTVRETDEDDHNEQLIIDLIIGSYGRRLKRRAG
GLNCRNPVQGEADSPEDSNLSAYIYAIKKKPLURRVYQILKILFRSNSDYG
NQTAEMWHTIIRKRGNGSSPACQKQLILNPKSGSGKGFYLFQKQAPVLITE
AEVOQDQTHHPQYAKFVTRRDLITRYSIVASGDGLFEVLINGMERMDWRRA
REDPLGII PCGSGNGLAKSVAHCEENPEYPIHLATLCKAGKCTPMDVVRVLA
RKHFPYVSI FLVSQNGMLTADIESERLTS ICAQRTLWAI KILGRSVKGVSYLL
GKGKKEPVEARELPASTAGI18S1PLNAGEFDLPDEEGBEAVLQGEQADATLS
LDRSVYRQHAUSHWGSMKRTAYISLGCPGSMNSRSMGISQRIEANAAGEFAEVPVTC
TIPPLQPLLSGDGWDGDFVMVHAAYTHLSSDVFPAESDQDGLLYLVLIRKG
VSRHQLLMFLNLNAGTLPIGEDPFIKVPCPRAFRIEPSSDGLIIVDGERVEYGI
OAEVNGPLINVTMTSQ"

```

```

Length      80.9%; Score 17.6; DB 3; Length 2629;
Similarity  90.5%; Pred.No. 8.7e-02;
19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1  CCGGCCAAAACCTCGCCTTA 21
   |||||
26 CCGGCATAAACCTCGCCTGA 2106

```

AC013897 46594 bp DNA linear HTG 16-NOV-1999
Drosophila melanogaster. *** SEQUENCING IN PROGRESS ***.

Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
 AC013897
 AC013897.1 GI:6437438
 HTG: HTGS_PHASE2.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Drosophila melanogaster
 Eukaryota: Metazoa: Arthropoda: Hexapoda: Insecta: Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 46594)

AUTHORS

Adams, M. and Venter, J.C.

TITLE

Direct Submission

JOURNAL

Rockville, MD, USA

COMMENT

This sequence was identified as CDM:10210473 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source

1..46594
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 46594;
Best Local Similarity 90.5%; Pred. No. 7.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 CCGGGCAAAACCTCGCCCTAA 21

|||||

DB

42642 CCGGGCATACCTCGCCCTGA 42662

RESULT 14

AC023718/c

LOCUS

164415 bp DNA linear INV 13-JUN-2002

DEFINITION

Drosophila melanogaster X BAC RP98-7P15 (Roswell Park Cancer
Institute Drosophila BAC Library) complete sequence.

ACCESSION

AC023718

VERSION

AC023718.4 GI:21397192

KEYWORDS

HTG.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster

REFERENCE

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

AUTHORS

Munry, D., Scherer, S., Adams, M.D., Holt, R.A., Evans, C.A.,
Gocayne, J.D., Taber, P., Williamson, A., Homs, F.H.,
Dugan-Rocha, S., Sodergren, E.S., Hodgson, A.H., Chen, R.C.,
Ayele, M., Scott, G.S., Worley, K.W., Amaratunga, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C.,
Busam, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K.,
Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E.,
Draper, H., Emery-Cohen, A., Ferreira, S., Garg, N.D.S., Houck, J.,
Hostin, D., Howland, T.J., Hume, J., Ibegwam, C., Jalali, M., Kovar, C.,
Liu, W., Mattei, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B.,
Nelson, K.A., Ndaesa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V.,
Scheeler, P., Shen, H., Strong, R., Tector, C., Wang, G., Williams, S.M.,
Xiang, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O.,
Wheeler, D., Weinstein, G., Gibbs, R. and Venter, J.C.

TITLE

Direct Submission

JOURNAL

Unpublished

AUTHORS

2 (bases 1 to 164415)

Worley, K.C., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaraturge, H.C., Are, J.R., Banks, T., Barbaria, J.,
Beaton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgall, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,

Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollans, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulseged, H., Lozardo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
Ogah, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Wleczyk, R., Wooden, S.,
Weinstock, G. and Gibbs, R.

TITLE

Unpublished

JOURNAL

Unpublished

AUTHORS

Worley, K.C.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

Submitted (17-FEB-2000)

AUTHORS

Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE

4 (bases 1 to 164415)

AUTHORS

Worley, K.C., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaraturge, H.C., Are, J.R., Banks, T., Barbaria, J.,
Beaton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgall, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garcia, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollans, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulseged, H., Lozardo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
Ogah, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Wleczyk, R., Wooden, S.,
Weinstock, G. and Gibbs, R.

Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Melgrim, J., Meneva, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataranan, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 17, 2003 this sequence version replaced gi17061039.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: wibr

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20088
Center clone name: 226_P_13

* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1312: contig of 1312 bp in length
* 1313 1412: gap of 100 bp
* 1413 10111: contig of 8699 bp in length
* 10112 10211: gap of 100 bp
* 10212 21544: contig of 11333 bp in length
* 21545 21644: gap of 100 bp
* 21645 74138: contig of 52494 bp in length
* 74139 74238: gap of 100 bp
* 74239 110959: contig of 36721 bp in length
* 110960 111059: gap of 100 bp
* 111060 166296: contig of 55237 bp in length.

FEATURES

SOURCE

1.166296
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-226P13"
/clone_lib="RPCI-24 Male Mouse BAC"
1.1312
/note="assembly_fragment"
1413.10111
/note="assembly_fragment"
10212.21544
/note="assembly_fragment"
21645.74138
/note="assembly_fragment"
74239.110959
/note="assembly_fragment"
111060.166296
/note="assembly_fragment"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 166296;
Best Local Similarity 90.5%; Pred. No. 7.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

2 CGGCAAAACCTCCGCTAAC 22

Db 89309 CGAGCAAAACCGCGCTAAC 89289

RESULT 16

AC117358

LOCUS

DEFINITION

AC117358

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

REVISION

AUTHORS

TITLES

JOURNAL

AC117358 229648 bp DNA linear HTG 19-NOV-2002
Rattus norvegicus clone CH230-307G20, WORKING DRAFT SEQUENCE.
AC117358
AC117358.5 GI:25085391
HTG; HTGS PHASR2; HTGS DRAFT; HTGS_FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Rattus norvegicus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 229648)
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chen, Y., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Loresuhewa, L., Louisege, H., Lozdo, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwaokeme, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfanckoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sheed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, A., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

AC117358 229648 bp DNA linear HTG 19-NOV-2002

Rattus norvegicus clone CH230-307G20, WORKING DRAFT SEQUENCE.

AC117358

AC117358.5 GI:25085391

HTG; HTGS PHASR2; HTGS DRAFT; HTGS_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 229648)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Y., Chen, Y.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,

Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Huiyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Loresuhewa, L., Louisege, H., Lozdo, R.J., Lu, X., Ma, J.,

Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,

Nwaokeme, O., Okwuonu, G., Olarpunsgoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfanckoch, C.,

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,

Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,

Sheed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,

Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,

Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,

Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,

Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K.,

Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, A., Yoon, V.,

Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von

Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,

Weinstock, G. and Gibbs, R.A.

Direct Submission

Unpublished

2 (bases 1 to 229648)

Worley, K.C.

Direct Submission

Submitted (10-APR-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 229648)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23811708.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence scaffolds are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GUNO
 Center clone name: CH230-307G20
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 22100 bases at least Q40
 Consensus quality: 22716 bases at least Q30
 Consensus quality: 223928 bases at least Q20
 Estimated insert size: 227432; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces
 is believed to be correct as given, however the sizes
 of the gaps between them are based on estimates that have
 been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

1 229648: contig of 229648 bp in length.
 Location/Qualifiers
 1..229648
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-307G20"
 1..2698
 /note="wgs end extension
 clone_end:Sp6"
 3433..4170
 /note="clone boundary
 clone_end:Sp6
 site:
 end sequence:BZ254694"
 183068..228499
 /note="clone boundary
 clone_end:17
 site:
 end sequence:BZ254693"
 228594..229648
 /note="wgs end extension
 clone_end:17"

ORIGIN
 Query Match 80.9%; Score 17.8; DB 2; Length 229648;

Best Local Similarity 90.5%; Pred. No. 7.5e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CCGGCGAAACCTCGGCTAA 21
 DB 210623 CCGGCGAAACATCCCTAA 210643
 RESULT 17
 AC097187
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-101L21, *** SEQUENCING IN PROGRESS
 ***, 3 unordered pieces.
 AC097187
 ACCESSION
 VERSION
 KEYWORDS
 HTG: HTGS_PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 242397)
 Muzny, D. Marie., Metzker, M. Lee., Abramzon, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, K., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregorgis, E., Gear, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamel, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hui, Y., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, J., London, P., Longacre, S., Lopez, J.,
 Lorusshewa, L., Louisedge, H., Lozado, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Nwankweli, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poinexter, A., Popovic, D., Primus, E., Pul, L.,
 Puato, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., S.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wiczyski, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstock, G. and Gibbs, R. A.
 Direct Submission
 Unpublished

TITLE
 JOURNAL

```

REFERENCE
AUTHORS      2 (bases 1 to 242397)
TITLE        Worley,K.C.
JOURNAL      Direct Submission
SUBMITTED    Submitted (12-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE    3 (bases 1 to 242397)
AUTHORS      Rat Genome Sequencing Consortium.
TITLE        Direct Submission
JOURNAL      Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT      On May 10, 2003 this sequence version replaced gi:23101628.
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
-----
Center project name: G1AU
Center clone name: CH230-101L21
----- Summary Statistics
-----
Assembly program: Atlas 3.0:
Consensus quality: 223422 bases at least Q40
Consensus quality: 227212 bases at least Q30
Consensus quality: 229858 bases at least Q20
Estimated insert size: 236640; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 3 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 237944: contig of 237944 bp in length
237945 238044: gap of unknown length
238045 239768: contig of 1724 bp in length
239769 239868: gap of unknown length
239869 242397: contig of 2529 bp in length.
----- Location/Qualifiers
source          1..242397
                 /organism="Rattus norvegicus"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:10116"
                 /clone="CH230-101L21"
misc_feature    1..1772
                 /note="wgs end extension
clone_end:Sp6"
                 5616..6387
misc_feature    /note="clone boundary
clone_end:Sp6"
                 site:EcoRI
                 end sequence: BH313336"
misc_feature    44761..46102
                 /note="wgs contig"
misc_feature    63661..64409

```

```

/note="clone boundary
clone_end:T7"
site:EcoRI
end_sequence: BH313334"
ORIGIN
Query Match      80.9%; Score 17.8; DB 2; Length 242397;
Best Local Similarity 90.5%; Pred. No. 7,5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCCGCCCTAA 21
   |||||
Db 239423 CCGGGCAAAACATCCCCCTAA 239443
RESULT 18
AC119383
LOCUS      254540 bp DNA linear HTG 19-SEP-2002
DEFINITION Rattus norvegicus clone CH230-221N23, *** SEQUENCING IN PROGRESS
            ***, 5 unordered pieces.
AC119383
AC119383.4 GI:23194991
VERSION    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS   Rattus norvegicus (Norway rat)
SOURCE     Rattus norvegicus
ORGANISM   Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
            Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 254540)
AUTHORS    Muzny,D.,Marie,, Metzker,M.,Lee,, Abranzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Dengson,S., Deramo,C., Ding,Y., Dinh,H., DiVya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M.,
            Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Loresuhwari,M., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwackelme,O., Okwuon,G., Olarnpungagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Richards,S., Riggs,F.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Snajds,D.,
            Speed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,K., Wei,X., White,F.,

```

Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 254540)

Worley,K.C.

Direct Submission

Submitted (26-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (19-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:21903160.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GWIK

Center clone name: CH230-221N23

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 230554 bases at least Q40

Consensus quality: 235111 bases at least Q30

Consensus quality: 237406 bases at least Q20

Estimated insert size: 258735; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

* NOTE: This sequence may represent more than one "clone"

* NOTE: This is a "working draft" sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 244381: contig of 244381 bp in length

* 244382 244481: gap of unknown length

* 24482 249870: contig of 5389 bp in length

* 249871 251359: gap of unknown length

* 249971 251359: contig of 1389 bp in length

* 251360 254459: gap of unknown length

* 251460 253223: contig of 1764 bp in length

* 253224 254540: gap of unknown length

* 253324 254540: contig of 1217 bp in length.

Location/Qualifiers

1. 254540

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-221N23"

1. 1113

/note="wgs contig"

244482. 255638

/note="wgs_contig"

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 254540;

Best Local Similarity 90.5%; Pred. No. 7.5e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCCCTAA 21

DB 25401 CCGGCAAAACATCCCTAA 25421

RESULT 19

AC115140

LOCUS

DEFINITION

AC115140.6 GI:23270043

HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 260335)

Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Ayodeji,A., Ayodeji,M., Baca,E, Baden,H., Balwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Evans,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,K., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Mallory,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K., Paoxleleleh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Frankoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Schefer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,

Weinstock, G. and Gibbs, R.A.
 Direct Submission
 Unpublished
 REFERENCE
 2 (bases 1 to 260335)
 Worley, K.C.
 Direct Submission
 Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 260335)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 23, 2002 this sequence version replaced gi:21736944.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequencing contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GNXU
 Center clone name: CH230-99N12
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 225253 bases at least Q40
 Consensus quality: 225495 bases at least Q30
 Consensus quality: 227152 bases at least Q20
 Estimated insert size: 239709; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 consists of 4 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 172711: contig of 172711 bp in length
 172811: gap of unknown length
 172812 191948: contig of 19137 bp in length
 191949 192048: gap of unknown length
 192049 258860: contig of 66812 bp in length
 258861 258960: gap of unknown length
 258961 260335: contig of 1375 bp in length.
 Location/Qualifiers
 1..260335
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-99N12"
 1..1294
 /note="wgs contig"
 175748..175884
 /note="wgs_contig"

FEATURES
 source
 misc_feature
 misc_feature
 ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 260335;
 Best Local Similarity 90.5%; Pred. No. 7.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 2 CGGCGAACCCTCGGCTTAC 22
 Db 223788 CGGAGAAAACCTCTCTTAC 223808
 RESULT 20
 AC133723 286789 bp DNA linear HTG 20-NOV-2002
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-107H7, *** SEQUENCING IN PROGRESS.
 ***, 8 unordered pieces.
 AC133723
 VERSION
 AC133723.2 GI:25138891
 KEYWORDS
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 SOURCE
 Rattus norvegicus (Norway rat)
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE
 1 (bases 1 to 286789)
 Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gregorogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Herrandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Huiy, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowls, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorusnewa, L., Louise, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Nwankwelu, O., Okwodu, G., Olapadehin, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
 Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wooden, H., Worley, K.,
 Williams, G., Willson, R., Wleczek, R., Wood, H., White, F.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.
 Direct Submission

JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 286789)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (18-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 286789)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 20, 2002 this sequence version replaced gi:23096637.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KCIF
 Center clone name: CH230-107H7
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 231208 bases at least Q40
 Consensus quality: 237333 bases at least Q30
 Consensus quality: 241696 bases at least Q20
 Estimated insert size: 241840; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: this is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 3691: contig of 3691 bp in length
 3692 3791: gap of unknown length
 3792 14444: contig of 10653 bp in length
 14445 14544: gap of unknown length
 14545 274434: contig of 259890 bp in length
 274435 274534: gap of unknown length
 274535 275544: contig of 1010 bp in length
 275545 277633: contig of 1989 bp in length
 277634 277733: gap of unknown length
 277734 279878: contig of 2145 bp in length
 279879 283039: gap of unknown length
 283040 283140: contig of 3661 bp in length
 283140 286789: contig of 3650 bp in length.
 Location/Qualifiers
 1..286789
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-107H7"

FEATURES
 source

misc_feature
 1..1219
 /note="wgs end extension
 clone end:Sp6"
 3792..5376
 /note="wgs end extension
 clone end:Sp6"
 14545..15732
 /note="wgs end extension
 clone end:Sp6"
 20525..21557
 /note="wgs end extension
 clone end:Sp6"
 25203..182322
 /note="clone boundary
 clone end:Sp6
 site:
 end sequence:BH360481"
 248024..249251
 /note="wgs contig"
 255512..258522
 /note="wgs contig"

ORIGIN
 Query Match 80.9%; Score 17.8; DB 2; Length 286789;
 Best Local Similarity 90.5%; Pred. No. 7.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCCGCTTAA 21
 |||||
 Db 151363 CCGGCAAAACATCCCTTAA 151383

RESULT 21
 AE003486/c
 LOCUS
 DEFINITION
 Drosophila melanogaster chromosome X section 38 of 74 of the
 complete sequence.
 ACCESSION
 AE003486 AE002593 AE014298
 VERSION
 AE003486.2 GI:228333082
 KEYWORDS
 SOURCE
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 328128)
 Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
 Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
 Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
 Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
 Wan,X.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L.,
 Abri,J.P., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,
 Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,B., Beasley,E.M.,
 Beeson,K.Y., Benos,P.V., Bereman,B.P., Bhandari,D., Bolshakov,S.,
 Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotter,P.,
 Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A.,
 Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,I.B.,
 Davis,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,
 Dinkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
 Ferreira,S., Fleischmann,W., Fowler,C., Gabriellian,A.E., Garg,N.S.,
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
 Wei,M.H., Iqbal,M., Kalish,W., Kallal,M., Kalush,F., Karpen,G.H., Ke,Z.,
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Milshina,N.V., Moberg,C.,
 McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Moberg,C.,
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,


```
/gene="dlgl"
/locus tag="CG1725"
/notes="dlgl gene product from transcript CG1725-RA"
/codon_start=1
/product="CG1725-PA"
/protein_id="AA48038.2"
/db_xref="GI:22833085"
/db_xref="FLYBASE:FBgn0001624"
/translation="MTTRKKRGGGGGGGGGFKKVSFLNLSLHKASSTKVGDDSW
LYEDLQERNGSLGRSAGCTGNPHGTDTSTYIKLJSGGAAADGRLSINDIIVS
INDVSVDVPHASADKAKNAGVKKLVHVRKGTATPAAGSAGDARDASAGPKV
IEIDLVRGKGLGSLGAGLGNQHPGDNIGYVTKMDGAAQVDGSLGSLKLIJVR
TNSEKLENVTHELAVATLKSITDKVLLIKTQHLTTSASGGGGGSLSSGQLSQS
QSOLATQSOSVHQHATPMVNSQSTGALNSQGTVDPSIPQAAAAVAANAAS
ASASVIASNNITNTTITVTATATASNSSSKLPPSLGANSISISINSNSNSNNIN
```

Query Match 80.9%; Score 17.8; DB 3; Length 328128;
Best Local Similarity 90.5%; Ref. No. 7.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCCGCTAA 21
|||||
Db 7598 CCGGGCATAACCTCCGCTGA 75968

RESULT 22
OSJN00127 163448 bp DNA linear PLN 14-NOV-2003
LOCUS Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJN0084K01,
DEFINITION complete sequence.
ACCESSION AL506999.3 GI:32489031
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1 Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., Li, Y., Zhu, J.,
Liu, Y., Hu, X., Jia, P., Zhang, Y., Zhao, Q., Ying, K., Yu, S., Fang, Y.,
Weng, Q., Zhang, L., Lu, Y., Mu, J., Lu, Y., Zhang, L. S., Yu, Z., Fan, D.,
Liu, X., Lu, T., Li, C., Wu, Y., Sun, T., Lei, H., Li, T., Hu, H., Guan, J.,
Wu, M., Zhang, R., Zhou, B., Chen, Z., Chen, L., Jin, Z., Wang, R.,
Yin, H., Cai, Z., Ren, S., Lv, G., Gu, W., Zhu, C., Tu, Y., Jia, J.,
Zhang, Y., Chen, J., Wang, H., Chen, X., Shao, C., Sun, Y., Hu, Q.,
Zhang, X., Zhang, J., Wang, L., Ding, C., Sheng, H., Gu, J., Chen, S.,
Ni, L., Zhu, F., Chen, W., Lan, L., Lai, Y., Cheng, Z., Gu, M., Jiang, J.,
Li, J., Hong, G., Xue, Y. and Han, B.
Sequence and analysis of rice chromosome 4
Nature 420 (6913), 316-320 (2002)

22337377
12447439

2
Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q., Hu, X.,
Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Weng, Q. J., Zhang, L.,
Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Zhang, Y. J., Lu, Y., Li, C.,
Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K., Zhou, B.,
Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P., Fu, G.,
Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F., Tu, Y. F.,
Jia, J., Yin, H. F., Zhang, Y., Cai, Z., Chen, J., Zhang, H., Chen, X. Y.,
Shao, C. Y., Sun, Y., Hu, Q. P., Zhang, X. L., Zhang, W., Wang, L. J.,
Ding, C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L., Zhu, F. H. and
Hong, G. F.
Direct Submission
Submitted (26-SEP-2001) Han Bin, National Center for Gene Research,
Chinese Academy of Sciences, 500# Cao Bao Road, Shanghai 200233,
China. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
bhan@ncgr.ac.cn
Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
clone: OSJN0084K01.
On Jul 9, 2003 this sequence version replaced gi:21912497.
Web site: http://www.ncgr.ac.cn

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL

REMARK
clone: OSJN0084K01.

COMMENT
On Jul 9, 2003 this sequence version replaced gi:21912497.
Web site: http://www.ncgr.ac.cn

----- Summary Statistics
Assembly program: phrap

This is a complete sequence.
Genes were identified by a combination of several methods: Gene
prediction programs including Fgenesh (<http://www.softberry.com/>),
genscan (<http://CCR-081.mit.edu/GENSCAN.html>), GeneMark-ES (Sean
Bddy, <http://genemark.biology.gatech.edu/GeneMark/>), tRNAscan-SE (Sean
Bddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the
complete sequence against NCBI none redundant protein database (nr)
(<http://ncbi.nlm.nih.gov/blast/db/>) and the EST database at NCGR.
Location/Qualifiers
1. 163448
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/variety="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="4"
/clone="OSJN0084K01"
/clone_lib="CUGI-OSJN0084K01.1"
/db_xref="GI:32489031"
/protein_id="CAE04830.1"
/codon_start=1
/db_xref="GI:32489032"
/translation="MTTRKKRGGGGGGGFKKVSFLNLSLHKASSTKVGDDSW
LYEDLQERNGSLGRSAGCTGNPHGTDTSTYIKLJSGGAAADGRLSINDIIVS
INDVSVDVPHASADKAKNAGVKKLVHVRKGTATPAAGSAGDARDASAGPKV
IEIDLVRGKGLGSLGAGLGNQHPGDNIGYVTKMDGAAQVDGSLGSLKLIJVR
TNSEKLENVTHELAVATLKSITDKVLLIKTQHLTTSASGGGGGSLSSGQLSQS
QSOLATQSOSVHQHATPMVNSQSTGALNSQGTVDPSIPQAAAAVAANAAS
ASASVIASNNITNTTITVTATATASNSSSKLPPSLGANSISISINSNSNSNNIN
EDBAAAAAAMASGERVLUKDE"
6124. .9960
/gene="OSJN0084K01.2"
/join(6124. .6288,6370. .6496,6568. .6698,7052. .7153,
7236. .7319,7403. .7630,7833. .8020,8122. .8226,8307. .8630,
9045. .9202,9298. .9404,9607. .9663,9751. .9960)
/gene="OSJN0084K01.2"
/codon_start=1
/protein_id="CAE04830.1"
/db_xref="GI:32489033"
/translation="MLQVEALAAVLASAAAARAAARAAVPLHDALGLVLAEDVRAP
DLPPFRASIKDGYAVVADSGPEVITESRAGDDALGVVTPGVAVVTCGPDP
GADAVQVDETLQAGAPGSKVRILVTRPQGDINRVGDIQKDSVVLKSEHIGP
AEGLATVGTITVKYRPTIAVSTGDELVEPTASLRGQIRDSNRLMLIAAIQ
HKGVLDLGVATTEESLKEHMDAALSDADILITSGVSMGDRDLVKFCLASMGKH
FKNRMPKGLTFAIEVITDSKSKVLAFLPGNVCVVCVFNLFVVPARRLSUG
WSNHLPRVHARLLHPLRGDPHREHRAVIRWPDGSGRPGYVAESVGGQASSEL
SKMSALLEVPLVGQMLEAGTSMQAILSDMTSPDKLPSTVNSSEKLGATVAT
STDLSQVPSQNTKVAITLTVSDTVSLGAGDRSGPRAISVNSSEKLGATVAT
AVPDDVEIKNLLVKSIDRSLITLGTGTGTPRDTVTEATKSVIEKAPGLTV
MLQESLKVTFPFAMLSRATAGIRGSTLIIINPGNPNAVASCMELLPLALKALKQIKGD
KREKHPRHVHAAAPVDQMDKSFRASSGRGSCRP"
complement(11086. .17771)
/gene="OSJN0084K01.3"
/complement(join(11086. .11128,11951. .12019,12589. .12645,
12738. .13527,13604. .13879,13950. .14016,14110. .14184,
14517. .14715,14826. .14975,15051. .15421,15969. .16061,
16145. .16402,17037. .17079,17667. .17771))
/gene="OSJN0084K01.3"
/codon_start=1
/protein_id="CAE04831.2"
/db_xref="GI:38346064"
/translation="MPMQPLAAGNMPHAAAOOPPPSYPPPPGOGVANHSAQ
SHGAGHNEEDWATKAWAAKSVTEHHIIOHAI SANRPHVNYGYDOYQCPAG
LPAPLPPIPPQSSNDLPQMTGQHQHRETNLYFDGGLPAPAKNFGSPSAQEVS
NYTSPGVNAMIQYPSSTQPLTASAMRDPFPQAPSSWHIAPSLQPHFGHGGSS
KIVADPSDQPLEFNSKAPDMVHRTANFNSNIIPAAIAASHDASAPASWIPSTT
VLFPQASVPPQAQMDPVSVAAPLFGAVGSYSYIPPAAGVGNVTFAPVDANTPNV
TESKPPVNNLRLELLKKKTSVGSVOHSTEFHSGKSEDAEKTLKSDQADSSVS
DSAKSLTDEDEDETEAARMAAMNOEIKVLTKVTDLLFDIATKVLNEDPS
AENEPTRSVSKPGLGSKPKVSAKVVPKPTNVSSADHSDGSLSPKGLLGL
ASVSDDEDEDEGDNEDKVPVSNLUSSETKAGARVGEKATDGEQGNHNRNVQN
ASLGEDRKNDEMSHRNSAELKQELRHDTHERFSDAKISQSDQGNHNRVQK
GYSQNGKVPSPNNVFNFSNMSQGHGKSGNEEDFVKPEKVRKSESSSKRYND

```

DDKSYNGINDKGSFKBKGSDRSVRHGADTRPRSRGNSKQDDAKGERKDYQKDV
EKSRDADRKEKDEKRSQTTKSSSHSRSPSARSRTKSSSHSRSPS
DEFSNDRKRLHSCSKSSMSPSPKSRNRVRSPSPHSHSRHSPSYSAERKKGRS
RTPVKRR"
/gene="OSJNBa0084K01.4"
complement (20406..21580)
/gene="OSJNBa0084K01.4"
complement (join(20406..20487,20766..20841,20927..21093,
21291..21360,21443..21580))
/gene="OSJNBa0084K01.4"
/codon_start=1
/protein_id="CAE04832.1"
/db_xref="GI:32489035"
/translation="WATSSAYPPPPYYRKYDYKDPSSAPPEPPPPVDGYPQLGAT
YTDVPLPSDQGVQLYPKSPDIDFKELRLNREQLHLELADILVERPQYAR
RVDSLIFKNLHLLNSLRPHQARATLHMLENQIRKQAIEDIKORREEAQKLLG
ESLLILDGNQPSLPAN"
23370..24831
/gene="OSJNBa0084K01.5"
join(23370..23556,23787..23879,23963..24044,24798..24831)
/gene="OSJNBa0084K01.5"
/codon_start=1
/protein_id="CAE04833.2"
/db_xref="GI:38346065"
/translation="MDVVKAAQLSGRTLEKVVVHPLVLLSIVDHYNRVARTRKRVVG
VLIGTSRSGSVDTNSYAVPPEDDKDPRIWFLDHNYESFMSFKRINAKHEVVGWY
STGPKLRENDLDIHALFNYYVPQCPGDY"
27864..29507
/gene="OSJNBa0084K01.6"
join(27864..28234,29348..29507)
/gene="OSJNBa0084K01.6"
/codon_start=1
/protein_id="CAE04834.2"
/db_xref="GI:38346066"
/translation="WAFSFSFPFRRCGGGGGSPKTSASASAAAAAGNDEBEL
GVTPQLDLRTLSDFSFSALQOGNSHTLAISISPPLSRGLDRDVRACFSQG
WKLNSPAGRSQTPCSPKPYELRAIQEKVRMETENGAKDVTIVEMQESKC
SRDSQTLPCDEESQDS"
complement (30293..31927)
/gene="OSJNBa0084K01.7"
complement (join(30293..30524,30814..30888,31535..31642,
31794..31927))
/gene="OSJNBa0084K01.7"
/codon_start=1
/protein_id="CAE04835.2"
/db_xref="GI:38346067"
/translation="WRKLPHEQKLLKTNFLDFKREKGRDAIVTORYLLVERDDYK
KNGICLWQKLVNLIKMDPRDPYRAEMTDLKLYNMGVITKSLTCEKLSVS
SFCCRRLATVVKLFAEHLKAVYIIQGGHVRVGPETVTPAPFLVTRNMEDFITWVD
SKIKRKQVQEVNGLDDEDFDAMA"
complement (35553..35984)
/gene="OSJNBa0084K01.8"
complement (35553..35984)
/gene="OSJNBa0084K01.8"
/codon_start=1
/protein_id="CAE04836.2"
/db_xref="GI:38346068"
/translation="MAKSKIRYIWLQRLTRWRRAAARAAAEAVPAGHVAVCVG
ASRFPVRAAHNHPVRELLRQAEYGFPGGAYCGPIALPCDEGLFEHLRLSSP
SSARFVLEIDIQSGALSCCAAGDALPLRLGIATDKAVN"
complement (50570..51031)
/gene="OSJNBa0084K01.9"
complement (50570..51031)
/gene="OSJNBa0084K01.9"
/codon_start=1
/protein_id="CAE04837.2"
/db_xref="GI:38346069"
/translation="MCKVITIQSLWLRLTRVRRAAARPAAGSKMERGDVAPGAGH
VAVRVGGGDEARFVVRVRELLRQAEYGFPGGAYCGPIALPCDEGLFEHLRLSSP
DVLRRVSDERHDLAFCRPAATCGVTKRAESSSSSRAPLQGMDSLVW"
complement (55313..57999)
/gene="OSJNBa0084K01.10"
complement (join(55313..55654,55732..56160,57637..57999))
/gene="OSJNBa0084K01.10"

```

```

/codon_start=1
/protein_id="CAE04838.2"
/db_xref="GI:38346070"
/translation="MAPVATITFLPTASNEATLPSVRDEDERPRVAVNQFSDAVPVI
SLQIGZARAEIRARVAGACEEWGIFQVVDHGVADGLVADMARLARDFFALPPDKL
RFMSGKKGGFIVSSHQEAVKDWEIVTFYSYFKSRDYSRWPKFAGWRAVVEQ
YSRLMGLACKGLVLSSEAGVDMADACVMDQKVYVFKPCQPDLTGLKRRH
TDFGTTITLLQDLVGGLOATRDAGTWTITVQPIPGSFVNVNLGSHAYLSNGRFRKNADH
Query Match 79.1% Score 17.4; DB 8; Length 163448;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCCGCCT 19
||| ||||| ||||| |||||
Db 27924 CCGAGCAAAACCTCCGCCT 27942
RESULT 23
BX510331 176859 bp DNA linear HTG 24-SEP-2003
LOCUS Best local clone CH211-42B22, WORKING DRAFT SEQUENCE, 3 unordered
DEFINITION Pieces.
ACCESSION BX510331
VERSION BX510331.5 GI:35209544
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 176859)
McLaren,S.
Direct Submission
Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 24, 2003 this sequence version replaced gi:31559357.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zC42B22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 175562 bases at least Q40
Consensus quality: 175799 bases at least Q30
Consensus quality: 175979 bases at least Q20
Insert size: 176659; sum-of-contigs
Quality score: 187706; 3.2% error; agarose-fp
Quality coverage: 8.99x in Q20 bases; sum-of-contigs Quality
coverage: 8.61x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* * 60644: contig of 60644 bp in length
* * 60745: gap of 100 bp
* * 60745 128253: contig of 67509 bp in length
* * 128254 128353: gap of 100 bp
* * 128354 176859: contig of 48506 bp in length.
* Location/Qualifiers
* 1..176859
* /organism="Danio rerio"
* /mol_type="genomic DNA"
FEATURES
source

```

```

/db xref="taxon:7955"
/clon="CH211-42B22"
/clone_lib="CHORI-211"
1..60644
/notes="assembly_fragment:02137"
fragment_chain:1
clone_end:T7
vector_side:left"
60745..128253
/notes="assembly_fragment:01319"
fragment_chain:1"
128354..176859
/notes="assembly_fragment:01889"
clone_end:SP6
vector_side:right"

ORIGIN
Query Match      79.18; Score 17.4; DB 2; Length 176859;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCGGCT 19
Db 54417 CTGGGCAAAACCTCGGCT 54435

RESULT 24
AC095761/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-9015, *** SEQUENCING IN PROGRESS ***,
9 unordered pieces.
AC095761
VERSION
AC095761.4 GI:22773150
KEYWORDS
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 325612)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Detamo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,W.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunarane,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huylk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuwa,L., Louised,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,K., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,

```

```

Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S.J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Snares, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausen, A., Weiss, R.A., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 325612)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 325612)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 10, 2002 this sequence version replaced gi:21722773.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDSK
Center clone name: CH230-9015
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 219952 bases at least Q40
Consensus quality: 225538 bases at least Q30
Consensus quality: 229145 bases at least Q20
Estimated insert size: 231139; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 286218: contig of 286218 bp in length
* 286219 286318: gap of unknown length
* 286319 293688: contig of 7370 bp in length
* 293689 293788: gap of unknown length

```

```

* 293789 294953: contig of 1165 bp in length
* 294954 295053: gap of unknown length
* 295054 296119: contig of 1066 bp in length
* 296120 296220: gap of unknown length
* 296220 297715: contig of 1496 bp in length
* 297716 297815: gap of unknown length
* 297816 299694: contig of 1879 bp in length
* 299695 299794: gap of unknown length
* 299795 301330: contig of 1535 bp in length
* 301330 302965: gap of unknown length
* 302966 303066: contig of 1536 bp in length
* 303066 325612: contig of 22547 bp in length.

```

FEATURES

```

source
  1..325612
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-9015"
    633..1430
    /note="clone boundary"
    clone_end:Sp6
    site:EcoRI
    end_sequence:BH344284"
    656..1330
    /note="clone boundary"
    clone_end:Sp6
    site:EcoRI
    end_sequence:BH344284"
    274711..276476
    /note="wgs_end-extension"
    clone_end:Sp6"

misc_feature
  79.1%; Score 17.4; DB 2; Length 325612;
  Best Local Similarity 94.7%; Pred. No. 1.2e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN
  Query Match 79.1%; Score 17.4; DB 2; Length 325612;
  Best Local Similarity 94.7%; Pred. No. 1.2e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 CCGGGCAAAACCTCGCCT 19

DB 69074 CCGGGCAAAACCTCGCCT 69056

RESULT 25

```

SEU43356
LOCUS
  SEU43356 2171 bp DNA linear BCT 12-APR-2001
  Salmonella enterica isocitrate lyase (aceA) gene, partial cds,
  isocitrate dehydrogenase kinase/phosphatase (aceK) gene, complete
  cds.

```

ACCESSION U43356

VERSION U43356.1 GI:1151303

KEYWORDS

SOURCE

ORGANISM

Salmonella enterica

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Salmonella.

1 (bases 1 to 2171)

Nelson, K., Wang, F.S. and Selander, R.K.

Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ,

208 Mueller Lab, University Park, PA 16802, USA

2 (bases 1 to 2171)

Boyd, E.F., Wang, F.S., Whittam, T.S. and Selander, R.K.

Molecular genetic relationships of the salmonellae

Appl. Environ. Microbiol. 62 (3), 804-808 (1996)

97076912

PUBLISHED 8975610

3 (bases 1 to 2171)

Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.

Size and sequence polymorphism in the isocitrate dehydrogenase

kinase/phosphatase gene (aceK) and flanking regions in Salmonella

enterica and Escherichia coli

Genetics 147 (4), 1509-1520 (1997)

JOURNAL

MEDLINE 98072475

PUBMED 9409817

FEATURES

```

source
  1..2171
    /organism="Salmonella enterica"
    /mol_type="genomic DNA"
    /strain="S2980"
    /sub_species="IIiia"
    /db_xref="taxon:28901"
    1..213
    /gene="aceA"
    <1..213
    /gene="aceA"
    /codon_start=1
    /transl_table=11
    /product="isocitrate lyase"
    /protein_id="AAC43890.1"
    /db_xref="GI:1151304"
    /translation="LAHAYAQEGMKHYVEKVQPEFAAAKDGTYTFVSHQQEVGTGYF
    DKVTIIQGGASSVTALTGSTESQF"
    307..2091
    /gene="aceK"
    307..2091
    /gene="aceK"
    /codon_start=1
    /transl_table=11
    /product="isocitrate dehydrogenase kinase/phosphatase"
    /protein_id="AAC43891.1"
    /db_xref="GI:1151305"
    /translation="MSRGLELLIAQTILQGDQYGRLEVTSGAQORFEQADWHAQV
    QAKSRHLYDHVHGLVAVQLRCLTDGKNTDANFLRVKHYVTRLLPDYPRPEIAESF
    ENSVYCRLPDHSLETPERLPFSSQPERFRTPPLAKQDFPDHGWELLMLRLSDI
    PLRLPQWKSRIIRIIIAHLTEILGEDALPRCHLOVANELFYRNKAALVGLKRLTPDG
    TILFPLPHRTDEGLFVDTCTTAAESIVFGFARSYFMVYAPLPAALVBLREILP
    GKTTAELNIAICQKHAKTESYREYLCYLAESDEFIEAPGIRGMVLMVFTLPQFDRV
    FKTIIDKFAQKENGAAHVACYQLVKEHVRGVADTQEFENFVLDKRDQIDPALMA
    LQOEAPKTIIDLGEHIVIRHIVIRRMVPLNIWLEQVEGQQLRDAIEYGNALROLAA
    ANIFGDMLFKNFVTRHGRVYVDYDEICVMYTVNFRDIPARVPEDLASEPWSYV
    SPGDVPEEFHRLCADPRIGLFEEMHADLFRADYWRALQTRIKEGHVEDVYAYRR
    ORFSVRYDIDRLDXAPASGNVCRTA"

```

gene

CDS

ORIGIN

```

Query Match 78.2%; Score 17.2; DB 1; Length 2171;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 CCGGGCAAAACCTCGCCTAAC 22

DB 1132 CCGGGCAAAACCTCGCCTAAC 1153

RESULT 26

SEU43357

LOCUS

DEFINITION

Salmonella enterica isocitrate lyase (aceA) gene, partial cds,

isocitrate dehydrogenase kinase/phosphatase (aceK) gene, complete

cds.

U43357.1 GI:1151306

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Salmonella enterica

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Salmonella.

1 (bases 1 to 2171)

Nelson, K., Wang, F.S. and Selander, R.K.

Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ,

208 Mueller Lab, University Park, PA 16802, USA

2 (bases 1 to 2171)

Boyd, E.F., Wang, F.S., Whittam, T.S. and Selander, R.K.

Molecular genetic relationships of the salmonellae

Appl. Environ. Microbiol. 62 (3), 804-808 (1996)

JOURNAL

MEDLINE	9076912	LOCUS	SEU43350	DEFINITION	Salmonella enterica isocitrate dehydrogenase cds.	ACCESSION	U43350	VERSION	U43350.1	GI:1151205	KEYWORDS	Salmonella enterica	ORGANISM	Salmonella enterica	REFERENCE	1 (bases 1 to 2467)	
PUBMED	8975610	LOCUS	SEU43350	DEFINITION	Salmonella enterica isocitrate dehydrogenase cds.	ACCESSION	U43350	VERSION	U43350.1	GI:1151205	KEYWORDS	Salmonella enterica	ORGANISM	Salmonella enterica	REFERENCE	1 (bases 1 to 2467)	
AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.	AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.	AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.	AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.	AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.	AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.	AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.	AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.	AUTHORS	Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.
TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli	TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli	TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli	TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli	TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli	TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli	TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli	TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli	TITLE	Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphate gene (aceK) and flanking regions in Salmonella enterica and Escherichia coli
JOURNAL	Genetics 147 (4), 1509-1520 (1997)	JOURNAL	Genetics 147 (4), 1509-1520 (1997)	JOURNAL	Genetics 147 (4), 1509-1520 (1997)	JOURNAL	Genetics 147 (4), 1509-1520 (1997)	JOURNAL	Genetics 147 (4), 1509-1520 (1997)	JOURNAL	Genetics 147 (4), 1509-1520 (1997)	JOURNAL	Genetics 147 (4), 1509-1520 (1997)	JOURNAL	Genetics 147 (4), 1509-1520 (1997)	JOURNAL	Genetics 147 (4), 1509-1520 (1997)
MEDLINE	89072475	MEDLINE	89072475	MEDLINE	89072475	MEDLINE	89072475	MEDLINE	89072475	MEDLINE	89072475	MEDLINE	89072475	MEDLINE	89072475	MEDLINE	89072475
FEATURES	9409817	FEATURES	9409817	FEATURES	9409817	FEATURES	9409817	FEATURES	9409817	FEATURES	9409817	FEATURES	9409817	FEATURES	9409817	FEATURES	9409817
source		source		source		source		source		source		source		source		source	
gene	1. .2171	gene	1. .2171	gene	1. .2171	gene	1. .2171	gene	1. .2171	gene	1. .2171	gene	1. .2171	gene	1. .2171	gene	1. .2171
CDS	<1. .213	CDS	<1. .213	CDS	<1. .213	CDS	<1. .213	CDS	<1. .213	CDS	<1. .213	CDS	<1. .213	CDS	<1. .213	CDS	<1. .213
	/organism="Salmonella enterica"		/organism="Salmonella enterica"		/organism="Salmonella enterica"		/organism="Salmonella enterica"		/organism="Salmonella enterica"		/organism="Salmonella enterica"		/organism="Salmonella enterica"		/organism="Salmonella enterica"		/organism="Salmonella enterica"
	/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"		/mol_type="genomic DNA"
	/strain="S2983"		/strain="S2983"		/strain="S2983"		/strain="S2983"		/strain="S2983"		/strain="S2983"		/strain="S2983"		/strain="S2983"		/strain="S2983"
	/sub_species="IIIA"		/sub_species="IIIA"		/sub_species="IIIA"		/sub_species="IIIA"		/sub_species="IIIA"		/sub_species="IIIA"		/sub_species="IIIA"		/sub_species="IIIA"		/sub_species="IIIA"
	/db_xref="taxon:28901"		/db_xref="taxon:28901"		/db_xref="taxon:28901"		/db_xref="taxon:28901"		/db_xref="taxon:28901"		/db_xref="taxon:28901"		/db_xref="taxon:28901"		/db_xref="taxon:28901"		/db_xref="taxon:28901"
	1. .213		1. .213		1. .213		1. .213		1. .213		1. .213		1. .213		1. .213		1. .213
	/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"
	<1. .213		<1. .213		<1. .213		<1. .213		<1. .213		<1. .213		<1. .213		<1. .213		<1. .213
	/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"		/gene="aceA"
	/codon_start=1		/codon_start=1		/codon_start=1		/codon_start=1		/codon_start=1		/codon_start=1		/codon_start=1		/codon_start=1		/codon_start=1
	/transl_table=11		/transl_table=11		/transl_table=11		/transl_table=11		/transl_table=11		/transl_table=11		/transl_table=11		/transl_table=11		/transl_table=11
	/product="isocitrate lyase"		/product="isocitrate lyase"		/product="isocitrate lyase"		/product="isocitrate lyase"		/product="isocitrate lyase"		/product="isocitrate lyase"		/product="isocitrate lyase"		/product="isocitrate lyase"		/product="isocitrate lyase"
	/protein_id="AAC43892.1"		/protein_id="AAC43892.1"		/protein_id="AAC43892.1"		/protein_id="AAC43892.1"		/protein_id="AAC43892.								

Nelson,K., Wang,F.S. and Selander,R.K.
 Direct Submission
 Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ,
 208 Mueller Lab, University Park, PA 16802, USA
 2 (bases 1 to 2467)
 Boyd,E.F., Wang,F.S., Whittam,T.S. and Selander,R.K.
 Molecular genetic relationships of the salmonellae
 Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
 97076912
 8975610
 3 (bases 1 to 2467)
 Nelson,K., Wang,F.S., Boyd,E.F. and Selander,R.K.
 Size and sequence polymorphism in the isocitrate dehydrogenase
 kinase/phosphatase gene (aceK) and flanking regions in Salmonella
 enterica and Escherichia coli
 Genetics 147 (4), 1509-1520 (1997)
 98072475
 94096197
 Location/Qualifiers
 1. .2467
 /organism="Salmonella enterica"
 /mol_type="genomic DNA"
 /strain="S2378"
 /sub_species="IIIB"
 /db_xref="taxon:28901"
 1. .213
 /gene="aceA"
 <1. .213
 /gene="aceA"
 /codon_start=1
 /transl_table=11
 /product="isocitrate lyase"
 /protein_id="AAC43878.1"
 /db_xref="GI:1151286"
 /translation="LAHAYAQEGMKHYVEKVPQPEFAAAXDGYTFTVSHQQEYGTGYQ
 DKVTTTIQGGASSVTALTGSTESQ"
 314. .2065
 /gene="aceK"
 314. .2065
 /gene="aceK"
 /codon_start=1
 /transl_table=11
 /product="isocitrate dehydrogenase kinase/phosphatase"
 /protein_id="AAC43879.1"
 /db_xref="GI:1151287"
 /translation="MPCRLLELLIAQTILQGFDAQYGRFLVTSGAQORPEQADQWQVQ
 QAMKSRHLYDHVHGVLQCLITDGKTDADFLLRVKEHYHTRLLPDPYRFBIASF
 FNSYQRLDFHRSLYRFLTFSSQGRFRFTLPRLKDFDPDHGWESLLMLRLSD
 FLRLPWKMSRDIYRITIAHLETFITGEDAPRCHQVANELFYENKAALVWGLKITPDG
 TLPLPILHRTDGEGLFVDTCLTITAEASIVFGFASYPVMVAPLPAALVWELREILP
 KTKTALYNALGCKHAKTESPVEYLCLAESDEKFIETAPGIRGMWLVFTLPQDFRV
 FKILKQAPAPKMSAHNRACVQLVKEHDVRGMRMADTQEFENFLDKQQLIDPALMAL
 LRQEAPEKTIIDLGEQIVIRHLYERRWVPLNWLQEGQGLRDALIEEYGNALRQLAA
 ANTFPGDMLFNKFGVTRHGRVWFYDEICYMTVEVNFDPPIPPAPYDEDELAASEPWSV
 SPQGVFPPEGRFHWLCAADRPILGPLFEEMHADLFPRADYWRALQTRIKEGHIEDVYAYRR
 ORFGVYGAISSTANS"

ORIGIN

```

Query Match          78.2%; Score 17.2; DB 1; Length 2467;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1  CCGGGCAAAACCTCCGGCTAAC 22
        |||||
Db      1139 CCGGGCAAAACCCACCGCCGAGC 1160

RESULT 28
SEU43351
LOCUS
DEFINITION
Salmonella enterica isocitrate lyase (aceA) gene, partial cds,
isocitrate dehydrogenase kinase/phosphatase (aceK) gene, complete
cds.

```

RESULT 28	SEU43351	2467 bp	DNA	linear	BCT 12-APR-2001
LOCUS	SEU43351				
DEFINITION	Salmonella enterica isocitrate lyase (aceA) gene, partial cds, isocitrate dehydrogenase kinase/phosphatase (aceK) gene, complete cds.				

```

ACCESSION      U43351
VERSION        U43351.1  GI:1151288
LOCUS          Salmonella enterica
DEFINITION     Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
ACCESSION     U43351
VERSION       U43351.1  GI:1151288
KEYWORDS      Salmonella enterica
SOURCE        Salmonella enterica
ORGANISM      Salmonella enterica
REFERENCE     1 (bases 1 to 2467)
AUTHORS      Nelson, K., Wang, F.S. and Sclander, R.K.
TITLE        Direct Submission
JOURNAL       Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ,
              208 Mueller Lab, University Park, PA 16802, USA
REFERENCE     2 (bases 1 to 2467)
AUTHORS      Boyd, E.F., Wang, F.S., Whittam, T.S. and Sclander, R.K.
TITLE        Molecular genetic relationships of the salmonellae
JOURNAL       Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
MEDLINE      97076912
PUBMED       8975610
REFERENCE     3 (bases 1 to 2467)
AUTHORS      Nelson, K., Wang, F.S., Boyd, E.F. and Sclander, R.K.
TITLE        Size and sequence polymorphism in the isocitrate dehydrogenase
              kinase/phosphatase gene (aceK) and flanking regions in Salmonella
              enterica and Escherichia coli
JOURNAL       Genetics 147 (4), 1509-1520 (1997)
MEDLINE      98072475
PUBMED       9409817
FEATURES      source
              Location/Qualifiers
                1..2467
                /organism="Salmonella enterica"
                /mol_type="genomic DNA"
                /strain="S2979"
                /sub_species="IIb"
                /db_xref="taxon:28901"
                1..213
                /gene="aceA"
                <1..213
                /gene="aceA"
                /codon_start=1
                /transl_table=11
                /product="isocitrate lyase"
                /protein_id="AAC43880.1"
                /db_xref="GI:1151289"
                /translation="LAHAYAOEGMKHYVEKVVQPEFAAKDGYTFVSHQEQVGTGYF
                DKVTIIQGGASSVTALTGSTSESQF"
                314..2065
                /gene="aceK"
                314..2065
                /gene="aceK"
                /codon_start=1
                /transl_table=11
                /product="isocitrate dehydrogenase kinase/phosphatase"
                /protein_id="AAC43881.1"
                /db_xref="GI:1151290"
                /translation="MPRCLELLIAQTILQGFDAQYGRFLVTSQAQRFQADWQAVQ
                QAMSRHLYDHHVGLVVEQLRCITDGKSTDAFLRVKHYTRLPLDYPFRIASFF
                FNSVYCLRFHSRTIPERLFISSQPERRTIPRLAKDFFPDHGWESLLMRILGDL
                PLRLPWONKSDIRYIIAHLTFEGEDALPRCHVOVANELFYRNKAALVGLKLTDPG
                TLPLLPHTHDEGELFVDCLTTTAEASIVFGFARSYFMVYAPLPAALVENVLEILP
                KXTAELYNALCGCKHAKTSRYLYCLAESEKFIPEAGIRGMVNLVFTLFGDVRV
                FKIINDKFAPEKMSAAHVRACVYOLVKEHDVRGMADTQEFNFVLDKRDIDPALMAL
                LRQAPKIDTLGSEQIVIRHLYIERMVPLNIWLEQEGQQLRDALIEEYNALRQLAA
                ANIFPGMFLFNFGVTHGVVYDYDEICVTEVNFPRDIPPARYPEDELSESPWYSV
                SRCDFVFEFRHMLCADPRGPLPEEMHDLFRADYWRALQTRIKEGHIEDVYAVRRR
                QRFGVRYGAISSTANS"
ORIGIN
Query Match      78.2%; Score 17.2; DB 1; Length 2467;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CCGGGCAAAACCTCCGCTAAC 22
        |||||
        U43347
        U43347.1  GI:1151276
        Salmonella enterica

Db      1139  CCGGGCAAAACCTCCGCTAAC 1160
        |||||
        U43347
        U43347.1  GI:1151276
        Salmonella enterica

```

```

RESULT 29
SEU43346
LOCUS          Salmonella enterica
DEFINITION     Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
ACCESSION     U43346
VERSION       U43346.1  GI:1151274
KEYWORDS      Salmonella enterica
SOURCE        Salmonella enterica
ORGANISM      Salmonella enterica
REFERENCE     1 (bases 1 to 2468)
AUTHORS      Nelson, K., Wang, F.S. and Sclander, R.K.
TITLE        Direct Submission
JOURNAL       Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ,
              208 Mueller Lab, University Park, PA 16802, USA
REFERENCE     2 (bases 1 to 2468)
AUTHORS      Boyd, E.F., Wang, F.S., Whittam, T.S. and Sclander, R.K.
TITLE        Molecular genetic relationships of the salmonellae
JOURNAL       Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
MEDLINE      97076912
PUBMED       8975610
REFERENCE     3 (bases 1 to 2468)
AUTHORS      Nelson, K., Wang, F.S., Boyd, E.F. and Sclander, R.K.
TITLE        Size and sequence polymorphism in the isocitrate dehydrogenase
              kinase/phosphatase gene (aceK) and flanking regions in Salmonella
              enterica and Escherichia coli
JOURNAL       Genetics 147 (4), 1509-1520 (1997)
MEDLINE      98072475
PUBMED       9409817
FEATURES      source
              Location/Qualifiers
                1..2468
                /organism="Salmonella enterica"
                /mol_type="genomic DNA"
                /strain="S2985"
                /sub_species="II"
                /db_xref="taxon:28901"
                1..2066
                /gene="aceA"
                <1..213
                /gene="aceA"
                /codon_start=1
                /transl_table=11
                /product="isocitrate lyase"
                /protein_id="AAC43871.1"
                /db_xref="GI:1151275"
                /translation="LAHAYAOEGMKHYVEKVVQPEFAAKDGYTFVSHQEQVGTGYF
                DKVTIIQGGASSVTALTGSTSESQF"
                315..2066
                /gene="aceA"
ORIGIN
Query Match      78.2%; Score 17.2; DB 1; Length 2468;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  CCGGGCAAAACCTCCGCTAAC 22
        |||||
        U43347
        U43347.1  GI:1151276
        Salmonella enterica

Db      1140  CCGGGCAAAACCTCCGCTAAC 1161
        |||||
        U43347
        U43347.1  GI:1151276
        Salmonella enterica

```


ORGANISM Salmonella enterica
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 2468)
AUTHORS Nelson, K., Wang, F.S. and Selander, R.K.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1995) Kimberlyn Nelson, Biology, Penn State Univ,
208 Mueller Lab, University Park, PA 16802, USA
REFERENCE 2 (bases 1 to 2468)
AUTHORS Boyd, E.F., Wang, F.S., Whittam, T.S. and Selander, R.K.
TITLE Molecular genetic relationships of the salmonellae
JOURNAL Appl. Environ. Microbiol. 62 (3), 804-808 (1996)
MEDLINE 97076912
PUBMED 8975610
REFERENCE 3 (bases 1 to 2468)
AUTHORS Nelson, K., Wang, F.S., Boyd, E.F. and Selander, R.K.
TITLE Size and sequence polymorphism in the isocitrate dehydrogenase
kinase/phosphatase gene (aceK) and flanking regions in Salmonella
enterica and Escherichia coli
JOURNAL Genetics 147 (4), 1509-1520 (1997)
MEDLINE 98072475
PUBMED 9409817
FEATURES
Location/Qualifiers
1..2468
/organism="Salmonella enterica"
/mol_type="genomic DNA"
/strain="S2993"
/sub_species="II"
/db_xref="taxon:28901"
1..213
/gene="aceA"
<1..213
/gene="aceA"
/codon_start=1
/transl_table=11
/product="isocitrate lyase"
/protein_id="AAC43872.1"
/db_xref="GI:1151277"
315..2066
/translation="LAHAYAGGGMKHVKEVQPEFAAAKDGITFVSHQEQVGTGYF
DKVTIIQGGASSYVLTGTSTESQF"
315..2066
/gene="aceK"
315..2066
/gene="aceK"
/codon_start=1
/transl_table=11
/product="isocitrate dehydrogenase kinase/phosphatase"
/protein_id="AAC43873.1"
/db_xref="GI:1151278"
/translation="MPRGLELLIACTILQGFDAQYGRFLEVTSGAQQRFQADWHAQV
QAKSRILYVDHVLGVLEQLRCITDCKSTDAFLRVKHYTRLDPYPRFIAESF
FNSVYCLFHRSLTPRLPIFSSQPERRTIPRLAKDFPDHGWESILMLILSDL
PLRUPWQKGRDIRIILAHUTETLGEDALPRCHVQVANELFYRNKAANLVCKLTTPDG
TLPELPIHRTDEBELFVDTLTAEASIVFGFARSYFMVYAPLPALVWELREILP
KTTAELMYAIGCCOKHAKTESYREYLCYLAESDEKFTAEFGIRGMVNLVFTLPFDRV
PKIITKDFAPKSAARACVACVQLVKEHVRGMADTQBFENFVLDKQIDPALMAL
LRQAPBKITDLGFOIVICHLYIERRVVPLNMLEQVEGQQLRDAIEEYGNALRQLAA
ANIIFPGMLEFNFGVYVDYDEICVMTVEVNFDRDIPPARYPEDELASBPWYSV
SPGDVFFEEFRHNLCAADPRIGLPIFEENHADLFRADYWRALQTRIKEGHIEDVYVYRR
QRFSVRYGAISSTANSS"

ORIGIN
Query Match 78.2%; Score 17.2; DB 1; Length 2468;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 CCGGGCAAACTCGCCTAAC 22
Db 1140 CCGGGCAAAACACCGCGGAGC 1161
Search completed: June 20, 2004, 11:43:17
Job time : 621.376 secs

243 14.8 67.3 902 6 ABQ46431 Oligonucle
c 244 14.8 67.3 902 6 ABQ46430 Oligonucle
245 14.8 67.3 918 5 AAH66019 C glutami
c 246 14.8 67.3 979 6 ABQ50762 Oligonucle
247 14.8 67.3 979 6 ABQ50763 Oligonucle
c 248 14.8 67.3 1008 7 ACF69175 Phototab
c 249 14.8 67.3 1033 6 ABQ33359 Oligonucle
c 250 14.8 67.3 1033 6 ABQ33358 Oligonucle
c 251 14.8 67.3 1038 8 AAD29428 DNA encod
c 252 14.8 67.3 1047 4 AAF67916 Coryneb
253 14.8 67.3 1047 4 AAF71619 Coryneb
254 14.8 67.3 1077 6 AAS97197 Neisseria
255 14.8 67.3 1106 5 AAS93396 DNA encod
c 256 14.8 67.3 1143 6 AAK93714 DNA of ni
c 257 14.8 67.3 1148 6 ABQ50999 Oligonucle
c 258 14.8 67.3 1148 6 ABQ50998 Oligonucle
c 259 14.8 67.3 1152 7 ACA41112 Prokaryot
c 260 14.8 67.3 1185 7 ABZ40256 N. gonorr
c 261 14.8 67.3 1185 7 ABZ41367 N. gonorr
c 262 14.8 67.3 1218 4 AAS56067 Salmonell
c 263 14.8 67.3 1243 7 ACC45235 Human CAP
c 264 14.8 67.3 1275 5 AAH52209 Human AFP
c 265 14.8 67.3 1311 7 ACA51494 Prokaryot
c 266 14.8 67.3 1314 4 ABL112985 Drosophil
c 267 14.8 67.3 1341 2 AAG90492 Rhodococc
c 268 14.8 67.3 1359 7 ACA49129 Prokaryot
c 269 14.8 67.3 1371 6 AAS97262 Neisseria
c 270 14.8 67.3 1374 7 ACA41912 Prokaryot
c 271 14.8 67.3 1401 4 AAS56282 Salmonell
c 272 14.8 67.3 1401 7 ACA51954 Prokaryot
c 273 14.8 67.3 1438 4 ABL02543 Drosophil
c 274 14.8 67.3 1449 7 ACA24419 Prokaryot
c 275 14.8 67.3 1473 7 ABZ40363 N. gonorr
c 276 14.8 67.3 1483 5 AAH49355 Rhodoc
c 277 14.8 67.3 1506 7 ACF71932 Phototab
c 278 14.8 67.3 1531 4 AAH21843 Rhodococc
c 279 14.8 67.3 1548 4 ABL30045 Drosophil
c 280 14.8 67.3 1646 4 ABL116835 Drosophil
c 281 14.8 67.3 1660 4 ABL158832 Human pol
c 282 14.8 67.3 1660 8 ADB48813 Novel hum
c 283 14.8 67.3 1756 7 ACA35508 Prokaryot
c 284 14.8 67.3 2017 4 AAH60618 Human pol
c 285 14.8 67.3 2028 6 ABK76947 Bacillus
c 286 14.8 67.3 2050 9 ADB53454 Primary r
c 287 14.8 67.3 2140 2 AAV42979 Streptoco
c 288 14.8 67.3 2140 2 AAZ96289 S. pneumo
c 289 14.8 67.3 2768 4 ABL18106 Drosophil
c 290 14.8 67.3 3490 4 ABL112984 Drosophil
c 291 14.8 67.3 3566 4 ABL02542 Drosophil
c 292 14.8 67.3 3633 2 AAQ75818 Non-A Non
c 293 14.8 67.3 3714 4 ABL30044 Drosophil
c 294 14.8 67.3 3945 4 ABL116834 Drosophil
c 295 14.8 67.3 3987 3 AAB81772 N. mening
c 296 14.8 67.3 4539 4 ABL03472 Drosophil
c 297 14.8 67.3 5034 3 AAB81492 N. mening
c 298 14.8 67.3 5068 5 AAF32508 Human mal
c 299 14.8 67.3 5080 4 ABL03231 Drosophil
c 300 14.8 67.3 5230 7 AAD49464 Human ves

ALIGNMENTS

RESULT 1
AAI99682_32/c
Continuation (33 of 45) of AAI99682 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000

AAI99682_05 500001 610000
AAI99682_06 600001 710000
AAI99682_07 700001 810000
AAI99682_08 800001 910000
AAI99682_09 900001 1010000
AAI99682_10 1000001 1110000
AAI99682_11 1100001 1210000
AAI99682_12 1200001 1310000
AAI99682_13 1300001 1410000
AAI99682_14 1400001 1510000
AAI99682_15 1500001 1610000
AAI99682_16 1600001 1710000
AAI99682_17 1700001 1810000
AAI99682_18 1800001 1910000
AAI99682_19 1900001 2010000
AAI99682_20 2000001 2110000
AAI99682_21 2100001 2210000
AAI99682_22 2200001 2310000
AAI99682_23 2300001 2410000
AAI99682_24 2400001 2510000
AAI99682_25 2500001 2610000
AAI99682_26 2600001 2710000
AAI99682_27 2700001 2810000
AAI99682_28 2800001 2910000
AAI99682_29 2900001 3010000
AAI99682_30 3000001 3110000
AAI99682_31 3100001 3210000
AAI99682_32 3200001 3310000
AAI99682_33 3300001 3410000
AAI99682_34 3400001 3510000
AAI99682_35 3500001 3610000
AAI99682_36 3600001 3710000
AAI99682_37 3700001 3810000
AAI99682_38 3800001 3910000
AAI99682_39 3900001 4010000
AAI99682_40 4000001 4110000
AAI99682_41 4100001 4210000
AAI99682_42 4200001 4310000
AAI99682_43 4300001 4410000
AAI99682_44 4400001 4411529

Query Match 100.0%; Score 22; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTAAC 22

DB 39846 CCGGGCAAAACCTCCGCTAAC 39825

RESULT 2

AAI99682_32/c
Continuation (33 of 44) of AAI99683 from base 3200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000

WPI 1800001 1910000
 AAI199683 18 1900001 2010000
 AAI199683 19 2000001 2110000
 AAI199683 20 2100001 2210000
 AAI199683 21 2200001 2310000
 AAI199683 22 2300001 2410000
 AAI199683 23 2400001 2510000
 AAI199683 24 2500001 2610000
 AAI199683 25 2600001 2710000
 AAI199683 26 2700001 2810000
 AAI199683 27 2800001 2910000
 AAI199683 28 2900001 3010000
 AAI199683 29 3000001 3110000
 AAI199683 30 3100001 3210000
 AAI199683 31 3200001 3310000
 AAI199683 32 3300001 3410000
 AAI199683 33 3400001 3510000
 AAI199683 34 3500001 3610000
 AAI199683 35 3600001 3710000
 AAI199683 36 3700001 3810000
 AAI199683 37 3800001 3910000
 AAI199683 38 3900001 4010000
 AAI199683 39 4000001 4110000
 AAI199683 40 4100001 4210000
 AAI199683 41 4200001 4310000
 AAI199683 42 4300001 4403765
 AAI199683 43

Query Match 100.0%; Score 22; DB 4; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCGGGCAAAACCTCCGCTTAA 22
 DB 34122 CCGGGCAAAACCTCCGCTTAA 34101

RESULT 3
 AAH52060
 ID AAH52060 standard; DNA; 723 BP.
 AC AAH52060;
 AT
 DT 04-SEP-2001 (first entry)
 DE Mycobacterium tuberculosis potential drug target gene SEQ ID 114.
 XX Drug target; growth; organism viability; characterisation; ds.
 KW
 OS Mycobacterium tuberculosis.
 PN WO200135317-A1.
 PD 17-MAY-2001.
 PF 13-NOV-2000; 2000WO-US031152.
 PR 12-NOV-1999; 99US-0165086P.
 PR 12-NOV-1999; 99US-0165124P.
 PR 01-FEB-2000; 2000US-0179531P.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Eisenberg D, Rotstein SH, Marcotte EM;
 XX WPI; 2001-329193/34.
 DR P-PSDB; AAGS1209.
 XX Identifying nucleotide or polypeptide sequence for use as drug target,
 PT involves providing algorithm that analyzes a functional relationship
 PT between nucleotide or polypeptide sequences, and comparing the sequences.
 XX
 PS Disclosure; Page 128; 207pp; English.
 XX

CC This invention relates to a method for identifying a nucleotide or
 CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism
 XX
 SQ Sequence 723 BP; 137 A; 224 C; 240 G; 122 T; 0 U; 0 Other;
 Query Match 95.5%; Score 21; DB 4; Length 723;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCGGGCAAAACCTCCGCTTAA 21
 DB 703 CCGGGCAAAACCTCCGCTTAA 723

RESULT 4
 ABL03297/c
 ID ABL03297 standard; cDNA; 2020 BP.
 XX
 AC ABL03297;
 AT
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 4373.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US009231.
 PR 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR P-PSDB; ABB59194.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 PS Claim 1; SEQ ID NO 4373; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737,
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

SQ Sequence 2020 BP; 500 A; 516 C; 589 G; 415 T; 0 U; 0 Other;
Query Match 80.9%; Score 17.8; DB 4; Length 2020;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTAA 21
DB 1986 CCGGGCATAACCTCCGCTGA 1966

RESULT 5
ACF35859/c
ID ACF35859 standard; cDNA; 2629 BP.
XX AC ACF35859;
XX DT 06-NOV-2003 (first entry)
XX DE D. melanogaster sphingosine kinase (SK)1 cDNA.
XX KW sphingolipid; cytosolic; gene therapy; cancer; sphingolipidase; SPL;
XX KW sphingosine-1-phosphate lyase; sphingosine kinase; SK1; gene; ss.
XX OS Drosophila melanogaster.
XX XX
XX Key Location/Qualifiers
XX FT 235..2160
XX FT /*tag= a
XX FT /*product= "SK"
XX FT /*note= "sphingosine kinase (ABR82392)"
XX FT 688..2160
XX FT /*tag= b
XX FT /*note= "SK1 fragment (ABR82389)"
XX PN WO2003062390-A2.
XX PD 31-JUN-2003.
XX PF 17-JAN-2003; 2003WO-US001739.
XX PR 17-JAN-2002; 2002US-00053510.
XX PR 17-JAN-2002; 2002US-0349582P.
XX PA (CHIL-) CHILDRENS HOSPITAL & RES CENT AT OAKLAND.
XX PI Saba JD, Fyrt H;
XX DR WPI; 2003-663394/62.
XX DR P-PSDB; ABR82389, ABR82392.
XX PT Identifying an agent that modulates sphingolipid metabolism for treating
XX PT e.g., breast cancer by culturing a homozygous null mutant Drosophila
XX PT melanogaster in the absence and presence of a candidate agent.
XX PS Claim 4; Page 127-128; 93pp; English.
XX CC The invention relates to identifying an agent that modulates sphingolipid
XX CC metabolism. The method is useful for identifying an agent that modulates
XX CC sphingolipid metabolism for preparing a composition for treating a
XX CC disease associated with altered sphingolipid metabolism in a patient,
XX CC e.g., colon cancer, breast cancer, uterine cancer, stomach cancer,
XX CC ovarian cancer, lung cancer, kidney cancer, adenocarcinoma of the rectum
XX CC or hereditary sensory neuropathy type 1 or any one of the
XX CC sphingolipidases. The present sequence represents a D. melanogaster
XX CC sphingosine kinase (SK)1 cDNA (GI:21429173)
XX SQ Sequence 2629 BP; 721 A; 625 C; 679 G; 604 T; 0 U; 0 Other;
Query Match 80.9%; Score 17.8; DB 8; Length 2629;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTAA 21
DB 2126 CCGGGCATAACCTCCGCTGA 2106

RESULT 6
ABL03296/c
ID ABL03296 standard; cDNA; 4020 BP.
XX AC ABL03296;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 4370.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABR59193.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 4370; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signaling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABR59737-
XX CC ABR72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at fcp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 4020 BP; 1245 A; 857 C; 893 G; 1025 T; 0 U; 0 Other;
Query Match 80.9%; Score 17.8; DB 4; Length 4020;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCCGCTAA 21
DB 2986 CCGGGCATAACCTCCGCTGA 2966

RESULT 7
ABQ14264/c
ID ABQ14264 standard; DNA; 511 BP.
XX AC ABQ14264;
XX DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 855.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX OS Homo sapiens.
 XX PN WO200218632-A2.
 XX PD 07-MAR-2002.
 XX PF 01-SEP-2001; 2001WO-EP010074.
 XX PP 01-SEP-2000; 2000DE-01043826.
 XX PR 05-SEP-2000; 2000DE-01044543.
 XX PA (EPIC-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX DR Determining the degree of cytosine methylation in genomic DNA, useful for
 XX diagnosis and prognosis, comprises selective hybridization of amplicons
 XX from chemically treated DNA.
 XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX SQ Sequence 511 BP; 78 A; 48 C; 185 G; 200 T; 0 U; 0 Other;
 Query Match 79.1%; Score 17.4; DB 6; Length 511;
 Best Local Similarity 94.7%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 4 GGCAAAACCTCGGCTAAC 22
 DB 25 GACAAACCTCGGCTAAC 7
 RESULT 8
 ABQ14265
 ID ABQ14265 standard; DNA; 511 BP.
 XX AC ABQ14265;
 XX DT 12-JUL-2002 (first entry)
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 856.
 XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW SNP; cell differentiation; ds.

XX KW Gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX OS Homo sapiens.
 XX PN WO200218632-A2.
 XX PD 07-MAR-2002.
 XX PF 01-SEP-2001; 2001WO-EP010074.
 XX PP 01-SEP-2000; 2000DE-01043826.
 XX PR 05-SEP-2000; 2000DE-01044543.
 XX PA (EPIC-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX DR Determining the degree of cytosine methylation in genomic DNA, useful for
 XX diagnosis and prognosis, comprises selective hybridization of amplicons
 XX from chemically treated DNA.
 XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX SQ Sequence 511 BP; 200 A; 185 C; 48 G; 78 T; 0 U; 0 Other;
 Query Match 79.1%; Score 17.4; DB 6; Length 511;
 Best Local Similarity 94.7%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 4 GGCAAAACCTCGGCTAAC 22
 DB 487 GACAAACCTCGGCTAAC 505
 RESULT 9
 ABQ35568/C
 ID ABQ35568 standard; DNA; 647 BP.
 XX AC ABQ35568;
 XX DT 12-JUL-2002 (first entry)
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22159.
 XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX

OS Homo sapiens.
XX WO200218632-A2.
XX
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP010074.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX 05-SEP-2000; 2000DE-01044543.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNP's); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ54121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention
XX
XX Sequence 647 BP; 62 A; 85 C; 233 G; 264 T; 0 U; 3 Other;
SQ
Query Match 78.2%; Score 17.2; DB 6; Length 647;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 CCGGGCAAAACCTCGGCTTAAAC 22
Db 187 CCGGCTAAACCTAGCCTTAAAC 166
RESULT 10
ABQ35569
ID ABQ35569 standard; DNA; 647 BP.
XX
XX AC ABQ35569;
XX
XX 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 22160.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX
XX WO200218632-A2.
XX

XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP010074.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX 05-SEP-2000; 2000DE-01044543.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNP's); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ54121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention
XX
XX Sequence 647 BP; 264 A; 233 C; 85 G; 62 T; 0 U; 3 Other;
SQ
Query Match 78.2%; Score 17.2; DB 6; Length 647;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 CCGGGCAAAACCTCGGCTTAAAC 22
Db 461 CCGGCTAAACCTAGCCTTAAAC 482
RESULT 11
ABL33576/C
ID ABL33576 standard; DNA; 5925 BP.
XX
XX AC ABL33576;
XX
XX 26-MAR-2002 (first entry)
XX
XX DE Human immune system associated gene SEQ ID NO: 1549.
XX
XX KW Human; immune system disease; cytosine methylation; antiaesthetic;
XX antiarteriosclerotic; antianaemic; cytosine; cytosine; cytosine;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antineumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antineumatic; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX
XX Homo sapiens.
XX

PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 XX
 PF 02-JUL-2001; 2001WO-EP007537.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 XX WPI; 2002-130909/17.
 DR
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 XX
 PS Claim 1; SEQ ID NO 1549; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 5925 BP; 1510 A; 105 C; 1246 G; 3064 T; 0 U; 0 Other;
 Query Match 78.2%; Score 17.2; DB 6; Length 5925;
 Best Local Similarity 86.4%; Pred. No. 2.1e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 CCGGCGAAACCTCGCCTAAC 22
 DB 4911 CCGGCGAAACCTCGCCTAAC 4890
 RESULT 12
 AAA43296/c
 ID AAA43296 standard; cDNA; 364 BP.
 XX
 AC AAA43296;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Xenopus secreted expressed sequence tag SEQ ID NO:2036.
 XX
 KW Human; mouse; xenopus; rat; secreted expressed sequence tag; sEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
 KW antitumor; osteoprotective; neuroprotective; neurotropic; antipsoriatic;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
 KW autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
 KW infection; depression; psoriasis; ss.
 XX
 OS Xenopus sp.
 XX
 PN WO200021990-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US024205.

XX
 PR 15-OCT-1998; 98US-0104435P.
 XX
 PA (GEM) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M;
 XX
 DR WPI; 2000-317937/27.
 XX
 PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (sESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders.
 XX
 PS Claim 1; Page 585; 618pp; English.
 XX
 CC AAA41261 to AAA43419 represent specifically claimed secreted expressed
 CC sequence tags (sESTs), isolated from human, mouse, xenopus and rat tissue
 CC sources. The sESTs can have a range of activities depending on the
 CC tissues they were isolated from. The activities include: chemotactic;
 CC proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
 CC haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
 CC antifungal; antiviral; antidiabetic; antiasthmatic; vulnary; antitumor;
 CC osteoprotective; neuroprotective; neurotropic; antiparkinsonian; antipsoriatic;
 CC cerebroprotective; anticonvulsant; and antidepressant. The sESTs can be
 CC used for gene therapy and in vaccines. The sESTs are useful as probes for
 CC the identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the sESTs. Proteins encoded by the sESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA43420 to AAA43425 represent linker variants which are given
 CC in the exemplification of the present invention
 XX
 SQ Sequence 364 BP; 111 A; 70 C; 100 G; 83 T; 0 U; 0 Other;
 Query Match 76.4%; Score 16.8; DB 3; Length 364;
 Best Local Similarity 90.0%; Pred. No. 2.7e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 CCGGCGAAACCTCGCCTAAC 21
 DB 176 CCGTCAAAACCTCTGCCTAA 157
 RESULT 13
 ABQ30690/c
 ID ABQ30690 standard; DNA; 502 BP.
 XX
 AC ABQ30690;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17281.
 XX
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PF 01-SEP-2001; 2001WO-EP010074.

PA 01-SEP-2000; 2000DE-01043826.
 XX 05-SEP-2000; 2000DE-01044543.
 XX (EPIG-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX Determining the degree of cytosine methylation in genomic DNA, useful for
 XX diagnosis and prognosis, comprises selective hybridization of amplicons
 XX from chemically treated DNA.
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX This invention describes a novel method for determining the degree of
 XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 XX genomic sample of DNA. The sample is treated chemically to convert
 XX cytosine (C) but not methylated C, to uracil, then part of the genomic
 XX DNA that contains the target C is amplified to form a labeled amplicon.
 XX The amplicon is hybridised to two classes, each with at least one member,
 XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 XX degree of hybridisation to both classes is determined from the label on
 XX the amplicon. From the ratio of labels hybridised to the two classes of
 XX oligomers, the degree of methylation is calculated. The method is used:
 XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 XX and of a wide range of diseases, e.g. cancer, disorders of the central
 XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 XX particularly by detecting mutations or single nucleotide polymorphisms
 XX (SNP's); and (ii) for differentiation of cell or tissue types and for
 XX investigating cell differentiation. The method allows the methylation
 XX status of many C residues to be determined simultaneously. ABQ13410-
 XX ABQ54121 represent genomic DNA sequences used to illustrate the method
 XX for determining the degree of cytosine methylation described in the
 XX disclosure of the invention
 XX Sequence 502 BP; 105 A; 47 C; 139 G; 211 T; 0 U; 0 Other;
 SQ Query Match 76.4%; Score 16.8; DB 6; Length 502;
 Best Local Similarity 90.0%; Pred. No. 2.8e-02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 CGGGCAAAACCTCGCCTAA 21
 DB 216 CGCGTAAACCTCGCCTAA 197
 RESULT 14
 ABQ30691
 ID ABQ30691 standard; DNA; 502 BP.
 XX AC ABQ30691;
 XX 12-JUL-2002 (first entry)
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 17282.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 XX Homo sapiens.
 XX WO200218632-A2.
 XX 07-MAR-2002.
 XX 01-SEP-2001; 2001WO-EP010074.
 XX 01-SEP-2000; 2000DE-01043826.
 XX 05-SEP-2000; 2000DE-01044543.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;

PA (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX WPI; 2002-371829/40.
 XX Determining the degree of cytosine methylation in genomic DNA, useful for
 XX diagnosis and prognosis, comprises selective hybridization of amplicons
 XX from chemically treated DNA.
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX This invention describes a novel method for determining the degree of
 XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 XX genomic sample of DNA. The sample is treated chemically to convert
 XX cytosine (C) but not methylated C, to uracil, then part of the genomic
 XX DNA that contains the target C is amplified to form a labeled amplicon.
 XX The amplicon is hybridised to two classes, each with at least one member,
 XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 XX degree of hybridisation to both classes is determined from the label on
 XX the amplicon. From the ratio of labels hybridised to the two classes of
 XX oligomers, the degree of methylation is calculated. The method is used:
 XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 XX and of a wide range of diseases, e.g. cancer, disorders of the central
 XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 XX particularly by detecting mutations or single nucleotide polymorphisms
 XX (SNP's); and (ii) for differentiation of cell or tissue types and for
 XX investigating cell differentiation. The method allows the methylation
 XX status of many C residues to be determined simultaneously. ABQ13410-
 XX ABQ54121 represent genomic DNA sequences used to illustrate the method
 XX for determining the degree of cytosine methylation described in the
 XX disclosure of the invention
 XX Sequence 502 BP; 211 A; 139 C; 47 G; 105 T; 0 U; 0 Other;
 SQ Query Match 76.4%; Score 16.8; DB 6; Length 502;
 Best Local Similarity 90.0%; Pred. No. 2.8e-02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 2 CGGGCAAAACCTCGCCTAA 21
 DB 287 CGCGTAAACCTCGCCTAA 306
 RESULT 15
 ABQ18020/c
 ID ABQ18020 standard; DNA; 502 BP.
 XX AC ABQ18020;
 XX 12-JUL-2002 (first entry)
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 4611.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX drug; side effect; cancer; central nervous system; cardiovascular;
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 XX Homo sapiens.
 XX WO200218632-A2.
 XX 07-MAR-2002.
 XX 01-SEP-2001; 2001WO-EP010074.
 XX 01-SEP-2000; 2000DE-01043826.
 XX 05-SEP-2000; 2000DE-01044543.
 XX (EPIG-) EPIGENOMICS AG.
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.
 XX
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 502 BP; 96 A; 48 C; 139 G; 219 T; 0 U; 0 Other;
 XX
 Query Match 76.4%; Score 16.8; DB 6; Length 502;
 Best Local Similarity 90.0%; Pred. No. 2.8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 CGGGCAAAACCTCCGCTAA 21
 Db 308 CGCGTAAACCTCCGCTAA 289
 XX
 RESULT 16
 ABQ18021
 ID ABQ18021 standard; DNA; 502 BP.
 XX
 AC ABQ18021;
 XX
 XX
 DT 12-JUL-2002 (first entry)
 DE
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 4612.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200218632-A2.
 PN
 XX 07-MAR-2002.
 PD
 XX
 XX 01-SEP-2001; 2001WO-EP010074.
 DF
 XX
 XX 01-SEP-2000; 2000DE-01043826.
 FR
 XX 05-SEP-2000; 2000DE-01044543.
 FR
 XX (EFIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;
 PI
 XX WPI; 2002-371829/40.
 DR
 XX

PT
 PT
 PT
 XX
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 502 BP; 219 A; 139 C; 48 G; 96 T; 0 U; 0 Other;
 XX
 Query Match 76.4%; Score 16.8; DB 6; Length 502;
 Best Local Similarity 90.0%; Pred. No. 2.8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 CGGGCAAAACCTCCGCTAA 21
 Db 195 CGCGTAAACCTCCGCTAA 214
 XX
 RESULT 17
 ABN75186/c
 ID ABN75186 standard; cDNA; 536 BP.
 XX
 AC ABN75186;
 XX
 XX
 DT 08-JUL-2002 (first entry)
 DE
 DE Human ORF133 cDNA, SEQ ID NO:265.
 XX
 XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200190366-A2.
 PN
 XX 29-NOV-2001.
 PD
 XX 24-MAY-2001; 2001WO-US017076.
 PF
 XX 24-MAY-2000; 2000US-0206690P.
 PR
 XX (CURA-) CURAGEN CORP.
 PA

XX PI Leach MD, Shinkets RA;
XX WPI; 2002-106200/14.
DR P-PSDB; ABP31160.
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and disorders related to organ
PT transplantation.
XX Claim 1; Page 335; 2508pp; English.
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (Open reading frame) 1-4534, and sequences ABN75054-
CC AAN7587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, thrombolytic activity,
CC receptor/ligand, anti-inflammatory activity, tumour inhibition activity,
CC and anti-infective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as epilepsy and Alzheimer's disease,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases
XX SQ Sequence 536 BP; 109 A; 138 C; 142 G; 147 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 6; Length 536;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 GGGCAAAACCTCCGCTAAC 22
DB 196 GGGCAAAAGCTCGGCTTAC 177
RESULT 18
ABQ25017
ID ABQ25017 standard; DNA; 565 BP.
XX AC ABQ25017;
XX 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 11608.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX Homo sapiens.

XX OS Homo sapiens.
XX WO200218632-A2.
XX 07-MAR-2002.
XX 01-SEP-2001; 2001WO-EP010074.
XX 01-SEP-2000; 2000DE-01043826.
XX 05-SEP-2000; 2000DE-01044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used;
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNPs); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ4121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention
XX SQ Sequence 565 BP; 249 A; 158 C; 50 G; 108 T; 0 U; 0 Other;
Query Match 76.4%; Score 16.8; DB 6; Length 565;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CGGGCAAAACCTCCGCTAA 21
DB 258 CGGGTAAACCTCCGCTAA 277
RESULT 19
ABQ25016/c
ID ABQ25016 standard; DNA; 565 BP.
XX AC ABQ25016;
XX 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 11607.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX Homo sapiens.

PN WO200218632-A2.
 XX 07-MAR-2002.
 PF 01-SEP-2001; 2001WO-EP010074.
 XX 01-SEP-2000; 2000DE-01043826.
 PR 05-SEP-2000; 2000DE-01044543.
 XX (EPIG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI; 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful for
 PT diagnosis and prognosis, comprises selective hybridization of amplicons
 PT from chemically treated DNA.
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX
 SQ Sequence 565 BP; 108 A; 50 C; 158 G; 249 T; 0 U; 0 Other;
 Query Match 76.4%; Score 16.8; DB 6; Length 565;
 Best Local Similarity 90.0%; Pred. No. 2.8e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 CGGCGCAAAACCTCGGCTAA 21
 DB 308 CGGCGTAACCTCGGCTAA 289
 RESULT 20
 ID ACA57563 standard; cDNA; 771 BP.
 XX ACA57563;
 AC
 XX 10-JUN-2003 (first entry)
 DT
 XX Human adipocyte Selected Interacting domain, SID, cDNA #650.
 DE
 XX Human; ss: gene; prey; adipocyte; SID; selected interacting domain;
 KW anorectic; antidiabetic; protein-protein interaction; diabetes;
 KW yeast 2-hybrid assay; metabolic disorder; obesity.
 XX
 OS Homo sapiens.
 OS
 XX WO200286122-A2.
 PN
 XX 31-OCT-2002.
 PD
 XX

PF 14-MAR-2002; 2002WO-EP003768.
 XX
 PR 14-MAR-2001; 2001US-0275734P.
 XX (HYBR-) HYBRIGENICS.
 PA
 XX Legrain P, Daviet L;
 PI
 XX WPI; 2003-103412/09.
 DR P-PSDB; ABU71019.
 DR
 XX
 XX New complex between two interacting proteins in adipocyte cells, useful
 PT for identifying selected interacting domains that modulate protein
 PT interactions, or for preventing or treating metabolic disorders such as
 PT obesity or diabetes.
 XX
 PS Claim 7; Page 326; 382pp; English.
 XX
 CC The invention relates to a complex between two interacting proteins in
 CC adipocyte cells, given in the specification. The proteins are identified
 CC by selecting a bait protein from a known adipocyte marker and then
 CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by
 CC members of an adipocyte cDNA library. The proteins are designated SID
 CC (RTM) (selected interacting domains) proteins. Also included are a
 CC polynucleotide encoding a polypeptide in the adipocyte cells, a
 CC recombinant host cell expressing at least one of the interacting
 CC polypeptides of the complex, selecting a modulating compound in adipocyte
 CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
 CC sequences given in the specification (including its fragment or variant),
 CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
 CC given in the specification (including its fragment or variant), a vector
 CC comprising the SID (RTM) polynucleotide, a recombinant host cell
 CC comprising all or part of the data, listed in the specification.
 CC The complex, polypeptides, polynucleotides and compounds are useful for
 CC preventing or treating metabolic disorders such as obesity or diabetes.
 CC The polynucleotides are useful as probes or primers. The complex is
 CC particularly useful for identifying selected interacting domains (SID
 CC (RTM)) for screening drugs that modulate the protein interaction, thus
 CC exhibiting the therapeutic effect. The present sequence encodes a SID
 CC (prey) protein of the invention
 XX
 SQ Sequence 771 BP; 232 A; 173 C; 188 G; 178 T; 0 U; 0 Other;
 Query Match 74.5%; Score 16.4; DB 7; Length 771;
 Best Local Similarity 94.4%; Pred. No. 4.4e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 CGGCGCAAAACCTCGGCT 19
 DB 404 CGGCGCAAAACCTCCACCT 421
 RESULT 21
 ID AAS53792/C
 XX AAS53792 standard; DNA; 927 BP.
 AC
 XX AAS53792;
 AC
 XX 13-FEB-2002 (first entry)
 DT
 XX Helicobacter pylori DNA for cellular proliferation protein #246.
 DE
 XX Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
 KW antibacterial; drug design.
 KW
 XX Helicobacter pylori.
 OS
 XX WO200170955-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 21-MAR-2001; 2001WO-US009180.
 PF

XX 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 21-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX (ELIT-) ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
DR P-PSDB; AAU35933.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX Claim 27; SEQ ID NO 7429; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 927 BP; 314 A; 176 C; 196 G; 241 T; 0 U; 0 Other;
Query Match 74.5%; Score 16.4; DB 4; Length 927;
Best Local Similarity 94.4%; Pred. No. 4.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 GGCAAAACCTCGGCTAA 21
DB 249 GGCAAAACCTCGGCTAA 232
RESULT 22
ABZ38654
ID ABZ38654 standard; DNA; 1077 BP.
XX AC ABZ38654;
XX 07-MAR-2003 (first entry)
XX N. gonorrhoeae nucleotide sequence SEQ ID 1897.
XX Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX Neisseria gonorrhoeae.
XX WO200279243-A2.
XX 10-OCT-2002.
XX 12-FEB-2002; 2002WO-IB002069.
XX 12-FEB-2001; 2001GB-0000342A.

XX (CHIR-) CHIRON SPA.
XX Fontana MR, Piza M, Massignani V, Monaci E;
XX WPI; 2003-058415/05.
DR P-PSDB; ABP77684.
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection.
XX Disclosure; Page 325; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
CC molecules of the invention
XX SQ Sequence 1077 BP; 235 A; 366 C; 296 G; 180 T; 0 U; 0 Other;
Query Match 74.5%; Score 16.4; DB 7; Length 1077;
Best Local Similarity 94.4%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGGCC 18
DB 844 CTGGGCAAAACCTCGGCC 861
RESULT 23
ACA48845/C
ID ACA48845 standard; DNA; 1320 BP.
XX AC ACA48845;
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #30502.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Salmonella paratyphi.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR P-PSDB; ABU44975.
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 14; SEQ ID NO 36715; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: the sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1320 BP; 260 A; 364 C; 403 G; 293 T; 0 U; 0 Other;

Query Match 74.5%; Score 16.4; DB 7; Length 1320;
 Best Local Similarity 94.4%; Pred. No. 4.5e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCGAACAACCTCGCC 18
 DB 1237 CCGGCGAACAACCTCGCC 1220

RESULT 24
 ABV76014
 ID ABV76014 standard; cDNA; 3832 BP.

AC ABV76014;

DT 11-FEB-2003 (first entry)

DE Human EGF pathway substrate 8 cDNA.

KW EGF pathway substrate 8; EPS8; human; cancer; diagnosis; cytostatic;
 KW vaccine; gene therapy; adoptive immunotherapy; Gene; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH CDS 210..2687
 FT /*tag= a
 FT /product= "Human EPS8"

XX WO200208045-A2.

XX 17-OCT-2002.

XX 03-APR-2002; 2002WO-US010439.

XX 04-APR-2001; 2001US-0281566P.

PA (GENZ) GENZYME CORP.

XX Charles AN;
 PI
 XX WPI; 2003-075469/07.
 DR P-PSDB; ABP58031.
 DR
 XX Novel composition for inducing immune response in a subject, comprises an
 PT immunogenic ligand individually characterized by its ability to elicit an
 PT immune response against the same native ligand.
 XX
 PS Claim 17; page 72-73; 76pp; English.

XX The present sequence is that of cDNA encoding human EGF pathway substrate
 CC 8 (EPS8). The invention provides compounds and methods for the detection,
 CC diagnosis and therapy of EPS8-related cancers and malignancies, and for
 CC monitoring the progress of EPS8-related cancers and malignancies.
 CC Immunogenic ligands (see ABP58032-34) and polynucleotides encoding them
 CC (see ABV76015-17) are also provided. A claimed method for inducing an
 CC immune response in a subject involves delivering a composition comprising
 CC at least one of these immunogenic ligands. A claimed method of aiding in
 CC the diagnosis of a neoplastic condition, or susceptibility to a
 CC neoplastic condition, involves determining a level of EPS8 mRNA in a
 CC sample that is at least twice that in a normal or control sample.
 CC detection using a probe or primer based on the present sequence. The
 CC invention also provides compositions which are useful as components of
 CC anti-cancer vaccines and for expanding immune effector cells that are
 CC specific for cancers characterised by expression of EPS8 and which can be
 CC used in adoptive immunotherapy

XX SQ Sequence 3832 BP; 1171 A; 781 C; 840 G; 1040 T; 0 U; 0 Other;

Query Match 74.5%; Score 16.4; DB 7; Length 3832;
 Best Local Similarity 94.4%; Pred. No. 4.8e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCGGCGAACAACCTCGCCCT 19

DB 1164 CCGGCGAACAACCTCGCCCT 1191

RESULT 25
 ABL03639/c
 ID ABL03639 standard; cDNA; 3963 BP.

XX ABL03639;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 5399.

DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEXE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB59536.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.
XX
PS Claim 1; SEQ ID NO 5399; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3963 BP; 924 A; 1087 C; 1143 G; 809 T; 0 U; 0 Other;
Query Match 74.5%; Score 16.4; DB 4; Length 3963;
Best Local Similarity 94.4%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GGGCAAAACCTCCGCCTA 20
|||
DB 3243 GGGCAAAACCTCCGCCTA 3226
RESULT 26
ABL03638/C
ID ABL03638 standard; cDNA; 10644 BP.
XX
AC ABL03638;
XX
XX 26-MAR-2002 (first entry)
XX
XX *Drosophila melanogaster* expressed polynucleotide SEQ ID NO 5396.
XX
XX *Drosophila*; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX *Drosophila melanogaster*.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB59535.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from *Drosophila* and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 5396; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from *Drosophila*. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 10644 BP; 3004 A; 2372 C; 2437 G; 2831 T; 0 U; 0 Other;
Query Match 74.5%; Score 16.4; DB 4; Length 10644;
Best Local Similarity 94.4%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GGGCAAAACCTCCGCCTA 20
|||
DB 8924 GGGCAAAACCTCCGCCTA 8907
RESULT 27
AAA81476/C
ID AAA81476 standard; DNA; 56485 BP.
XX
XX AAA81476;
AC
XX
XX 04-DEC-2000 (first entry)
XX
XX N. meningitidis partial DNA sequence gnm_24 SEQ ID NO:24.
XX
XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
XX antigen; vaccine; diagnosis; infection; antibacterial; identification;
XX *Meningococcus B*; MenB; ds.
XX
XX *Neisseria meningitidis*.
XX
XX WO200022430-A2.
XX
XX 20-APR-2000.
XX
XX 08-OCT-1999; 99WO-US023573.
XX
XX 09-OCT-1998; 98US-0103794P.
XX
XX 30-APR-1999; 99US-0132068P.
XX
XX (CHIR) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
XX Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX Rappuoli R, Pizza M;
XX
XX WPI; 2000-318079/27.
XX
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be used
XX in the diagnosis and treatment of *N. meningitidis* infection and other
XX *Neisseria* infections, for example, *N. gonorrhoea*.
XX
XX Claim 7; Page 507-524; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic proteins
XX from *Neisseria* genomic sequences. AAA81453 to AAA82414 represent
XX specifically claimed *Neisseria meningitidis* genomic DNA sequences;
XX AAA81260 to AAA81303 and AAB25620 to AAB25663 represent *Neisseria* DNA
XX sequences and their corresponding proteins; AAA81254 to AAA81259 and
XX AAA81304 to AAA81321 represent PCR primers used in the isolation of
XX *Neisseria meningitidis* DNA sequences; and AAA81322 to AAA81452 represent
XX *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all
XX used in the exemplification of the present invention. The nucleic acid
XX sequences, protein sequences, and antibodies against them, can be used in
XX the manufacture of a composition. The composition can be used as a
XX medicament (or in the manufacture of a medicament) for treating,
XX preventing or diagnosing infection due to *Neisseria* bacteria. For
XX example, some of the identified proteins could be components of vaccines
XX against *Meningococcus B*; against all serotypes; and/or against all
XX pathogenic *Neisseriae*. Identification of sequences from the bacterium
XX will also facilitate production of biological probes, particularly
XX organism-specific probes. Attempts to make efficacious *Meningococcus B*
XX vaccines have failed mainly due to antigen tolerance. Multivalent
XX vaccines have also been tried but none have successfully overcome
XX antigenic variability. The provision of further, complete sequences may
XX provide an opportunity to identify secreted or surface exposed proteins

CC that may be presumed targets for the immune system and which are not
 CC antigenically variable or at least more conserved than other more
 CC variable regions
 XX

SQ Sequence 56485 BP; 12504 A; 14247 C; 16158 G; 13573 T; 0 U; 3 Other;

Query Match 74.5%; Score 16.4; DB 3; Length 56485;
 Best Local Similarity 94.4%; Pred. No. 5.7e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCC 18
 DB 35406 CTGGCAAAACCTCGCC 35389

RESULT 28
 AAA81489_4
 Continuation (5 of 9) of AAA81489 from base 400001 (N. meningitidis partial DNA sequence
 WP Sequence split into 9 fragments LOCUS AAA81489 Accession AAA81489

WP	Fragment Name	Begin	End
WP AAA81489_1	1	100001	110000
WP AAA81489_2	2	100001	210000
WP AAA81489_3	3	200001	310000
WP AAA81489_4	4	300001	410000
WP AAA81489_5	5	400001	510000
WP AAA81489_6	6	500001	610000
WP AAA81489_7	7	600001	710000
WP AAA81489_8	8	700001	810000
WP AAA81489_9	9	800001	837096

Query Match 74.5%; Score 16.4; DB 3; Length 110000;
 Best Local Similarity 94.4%; Pred. No. 5.8e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCC 18
 DB 71019 CTGGCAAAACCTCGCC 71036

RESULT 29
 AAF21612
 ID AAF21612 standard; DNA; 349980 BP.

XX AAF21612;

XX 13-MAR-2001 (first entry)

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KW ds.

OS Neisseria meningitidis.

XX WO200066791-A1.

XX 09-NOV-2000.

XX 08-MAR-2000; 2000WO-US005928.

XX 30-APR-1999; 99US-0132068P.

XX 08-OCT-1999; 99WO-US023573.

XX 28-FEB-2000; 2000GB-00004695.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC;

XX Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

XX Rappuoli R, Frazer CM, Grandi G;

XX WPI; 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections.
 XX

XX Claim 7; Appendix A; 692pp; English.

XX The present invention describes the full length genome of Neisseria
 CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
 CC represent fragments of the NMB genomic sequence, as the sequence was too
 CC long to go in a record on its own it was split into 8 sequences which
 CC overlap each other at the beginning and end of each sequence by 49980 bp
 CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
 CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
 CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
 CC given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
 CC primers which are used in the exemplification of the present invention.
 CC The NMB genome and fragments from it have antibacterial activity, and can
 CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
 CC and/or antibodies which binds to the proteins can be used in compositions
 CC for treating or preventing infection due to Neisserial bacteria or as a
 CC diagnostic reagent for detecting the presence of Neisserial bacteria or
 CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
 CC computer storage medium or computer databases can be used in a search to
 CC identify open reading frames (ORFs) or coding sequences within the NMB
 CC genome. The DNA sequences provide further opportunities to find antigenic
 CC or immunogenic proteins which are more effective in vaccines than the
 CC outer membrane proteins currently used

XX SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 U; 0 Other;

Query Match 74.5%; Score 16.4; DB 3; Length 349980;

Best Local Similarity 94.4%; Pred. No. 6.1e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCC 18

DB 106248 CTGGCAAAACCTCGCC 106265

RESULT 30

ABZ52802/c

ID ABZ52802 standard; CDNA; 461 BP.

XX ABZ52802;

XX 28-MAR-2003 (first entry)

XX Aspergillus oryzae polynucleotide SEQ ID NO 1915.

XX Aspergillus oryzae; fermentation; fungus; industrial; EST;
 KW expressed sequence tag; gene; ss.

XX Aspergillus oryzae.

XX WO200279476-A1.

XX 10-OCT-2002.

XX 22-MAR-2002; 2002WO-IB000890.

XX 30-MAR-2001; 2001JP-00098371.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

PA (NARE-) NAT RES INST BREWING.

PA (NORQ) NAT FOOD RES INST MIN AGRIC.

PI Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;

PI Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;

XX WPI; 2003-046817/04.

XX Detection of expression of specific Aspergillus genes for monitoring the
 PT fermentation and growth conditions of the fungus, using DNA probes.

XX Claim 1; SEQ ID NO 1915; 48pp + Sequence Listing; Japanese.
 PS
 XX
 CC The invention relates to a polynucleotide having any of 6006 specific
 CC sequences (AB250888-AB256933), which are expressed by a fungus under
 CC specific culture conditions including one or more of eutrophic,
 CC oligotrophic, solid, early germination, alkaline, high temperature, low
 CC temperature or maltose culture or polynucleotides stringently hybridising
 CC to these sequences. The polynucleotides are useful for monitoring the
 CC progress of fermentation and the growth conditions of a fungus, rial
 CC especially of Aspergillus oryzae which is widely used in industrial
 CC fermentation. Also monitoring for fungal contamination. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 461 BP; 105 A; 134 C; 110 G; 112 T; 0 U; 0 Other;
 Query Match 73.6%; Score 16.2; DB 7; Length 461;
 Best Local Similarity 85.7%; Pred. No. 5.3e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 CGGGCAAAACCTCGCCTAAC 22
 Db 96 CGGGCAAAACCTCGCCTAAC 76

Search completed: June 20, 2004, 10:17:20
 Job time : 144.538 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1028.06 Seconds
(without alignments)
639.034 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22
Sequence: 1 cogggcaaacctccgcctaac 22

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026579

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :

EST.*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmd:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssi:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	18.8	85.5	246	12	BJ206364
C 2	18.8	85.5	355	12	BJ176181
C 3	18.8	85.5	365	12	BJ201391
C 4	18.8	85.5	390	12	BJ185747

C 5	18.8	85.5	396	12	BJ185396
C 6	18.8	85.5	406	12	BJ175801
C 7	18.8	85.5	420	12	BJ181492
C 8	18.8	85.5	427	12	BJ183162
C 9	18.8	85.5	503	12	BJ180094
C 10	18.8	85.5	564	12	BJ198230
C 11	18.8	85.5	586	12	BJ183579
C 12	18.8	85.5	666	28	BJ375652
C 13	17.8	80.9	326	10	BE226036
C 14	17.8	80.9	394	12	EM944427
C 15	17.8	80.9	399	10	BF546805
C 16	17.8	80.9	412	10	BE226096
C 17	17.8	80.9	441	9	AI465361
C 18	17.8	80.9	458	9	AI265329
C 19	17.8	80.9	467	10	BE863645
C 20	17.8	80.9	484	9	AA584005
C 21	17.8	80.9	487	14	CD807506
C 22	17.8	80.9	500	14	R48986
C 23	17.8	80.9	510	12	BM950311
C 24	17.8	80.9	532	9	AI598373
C 25	17.8	80.9	606	13	BU059049
C 26	17.8	80.9	658	14	CK134467
C 27	17.8	80.9	697	28	BZ253738
C 28	17.8	80.9	714	13	BQ863377
C 29	17.8	80.9	812	13	BU847656
C 30	17.8	80.9	989	9	AV154279
C 31	17.8	80.9	1101	29	CNS00017
C 32	17.8	80.9	1201	29	CNS016FG
C 33	17.4	79.1	324	28	AQ077340
C 34	17.4	79.1	557	14	CB488692
C 35	17.4	79.1	1036	28	CC268965
C 36	17.2	78.2	427	10	AW596752
C 37	17.2	78.2	431	13	BU577375
C 38	17.2	78.2	513	9	A1736328
C 39	17.2	78.2	513	10	AW836648
C 40	17.2	78.2	533	14	CB352999
C 41	17.2	78.2	537	12	BI974766
C 42	17.2	78.2	559	10	AW655743
C 43	17.2	78.2	628	28	BH065920
C 44	17.2	78.2	782	13	BX880282
C 45	17.2	78.2	790	14	CF455958
C 46	17.2	78.2	916	13	EX414416
C 47	17.2	78.2	959	13	BQ430948
C 48	16.8	76.4	288	12	BI071295
C 49	16.8	76.4	389	12	BG799408
C 50	16.8	76.4	417	13	BY083094
C 51	16.8	76.4	428	10	BF601955
C 52	16.8	76.4	457	9	AV616386
C 53	16.8	76.4	493	10	AW768210
C 54	16.8	76.4	497	29	CG898480
C 55	16.8	76.4	516	12	BI138891
C 56	16.8	76.4	535	13	BU825910
C 57	16.8	76.4	544	14	CB433177
C 58	16.8	76.4	548	9	AV749560
C 59	16.8	76.4	570	14	CF447487
C 60	16.8	76.4	612	29	CG898304
C 61	16.8	76.4	625	29	CG898411
C 62	16.8	76.4	629	14	CD895297
C 63	16.8	76.4	652	29	CG896231
C 64	16.8	76.4	692	29	CG898913
C 65	16.8	76.4	699	29	CG895869
C 66	16.8	76.4	700	29	CG895984
C 67	16.8	76.4	712	29	CG896092
C 68	16.8	76.4	720	13	EX846241
C 69	16.8	76.4	723	29	CG820944
C 70	16.8	76.4	737	29	CG898879
C 71	16.8	76.4	747	14	CF440475
C 72	16.8	76.4	750	28	BZ045990
C 73	16.8	76.4	754	13	BU260771
C 74	16.8	76.4	754	29	CG896553
C 75	16.8	76.4	778	14	CB288874
C 76	16.8	76.4	779	14	CF455372
C 77	16.8	76.4	780	29	CG896489

BJ185396	BJ185396
BJ175801	BJ175801
BJ181492	BJ181492
BJ183162	BJ183162
BJ180094	BJ180094
BJ198230	BJ198230
BJ183579	BJ183579
BJ375652	AG-ND-178
BE226036	ial3a09.Y
EM944427	UI-M-EHOP
BF546805	UI-R-C2P-
BE226096	ial4a08.Y
AI465361	mu75g03.X
AI265329	uk01g08.Y
BE863645	UI-M-BHO-
AA584005	nn65a10.S
CD807506	UI-M-GWO-
R48986	YG65b01.S1
BM950311	UI-M-EHOP
AI598373	EST250076
BU059049	UI-M-PRO-
CK134467	RE84552.3
BZ253738	CH230.306
BQ863377	QG233.12.
BU847656	AGENCOURT
AV154279	AV154279
AL058203	Drosophil
AL105678	Drosophil
AQ077340	CIT-HSP-2
CB488692	omykrtfho
CC268965	CH261.58L
AW596752	sj16c09.Y
BU577375	sa69h08.
A1736328	sb27a10.Y
AW836648	sm04h07.Y
CB352999	ZF001-P00
BI974766	sa172f10.
AW655743	106969.MA
BH065920	RPGT-24-3
BX880282	BX880282
CF455958	EST682303
EX414416	EX414416
BQ430948	AGENCOURT
BI071295	C054P79U
BG799408	fm96f03.Y
BY083094	BY083094
BF601955	267067.MA
AV616386	AV616386
AW768210	da70g03.Y
CG898480	pastbac07
BI138891	FL18P50Y
BU825910	UK1.14TB06
CB433177	609500.MA
AV749560	AV749560
CF447487	EST683832
CG898304	pastbac06
CG898411	pastbac07
CD895297	GL174.001H
CG896231	pastbac01
CG898913	pastbac08
CG895869	pastbac09
CG895984	pastbac00
CG896092	pastbac00
EX846241	EX846241
CG820944	FCFP3A10
CG898879	pastbac08
CF440475	EST676820
CF455372	AGENCOURT
CG896489	pastbac02

C 78	16.8	76.4	794	14	CF290814	AGENCYCOURT
C 79	16.8	76.4	847	14	CD301301	AGENCYCOURT
C 80	16.8	76.4	869	14	CK156714	FOYDRS9772
C 81	16.8	76.4	876	29	CG822538	SOYDRS9772
C 82	16.8	76.4	901	13	CG954414	AGENCYCOURT
C 83	16.8	76.4	916	13	CG954414	AGENCYCOURT
C 84	16.8	76.4	924	13	CG954414	AGENCYCOURT
C 85	16.8	76.4	930	13	CG954414	AGENCYCOURT
C 86	16.8	76.4	936	13	CG954414	AGENCYCOURT
C 87	16.8	76.4	942	13	CG954414	AGENCYCOURT
C 88	16.8	76.4	948	13	CG954414	AGENCYCOURT
C 89	16.8	76.4	954	13	CG954414	AGENCYCOURT
C 90	16.8	76.4	960	13	CG954414	AGENCYCOURT
C 91	16.8	76.4	966	13	CG954414	AGENCYCOURT
C 92	16.8	76.4	972	13	CG954414	AGENCYCOURT
C 93	16.8	76.4	978	13	CG954414	AGENCYCOURT
C 94	16.8	76.4	984	13	CG954414	AGENCYCOURT
C 95	16.8	76.4	990	13	CG954414	AGENCYCOURT
C 96	16.8	76.4	996	13	CG954414	AGENCYCOURT
C 97	16.8	76.4	1002	13	CG954414	AGENCYCOURT
C 98	16.8	76.4	1008	13	CG954414	AGENCYCOURT
C 99	16.8	76.4	1014	13	CG954414	AGENCYCOURT
C 100	16.8	76.4	1020	13	CG954414	AGENCYCOURT
C 101	16.8	76.4	1026	13	CG954414	AGENCYCOURT
C 102	16.8	76.4	1032	13	CG954414	AGENCYCOURT
C 103	16.8	76.4	1038	13	CG954414	AGENCYCOURT
C 104	16.8	76.4	1044	13	CG954414	AGENCYCOURT
C 105	16.8	76.4	1050	13	CG954414	AGENCYCOURT
C 106	16.8	76.4	1056	13	CG954414	AGENCYCOURT
C 107	16.8	76.4	1062	13	CG954414	AGENCYCOURT
C 108	16.8	76.4	1068	13	CG954414	AGENCYCOURT
C 109	16.8	76.4	1074	13	CG954414	AGENCYCOURT
C 110	16.8	76.4	1080	13	CG954414	AGENCYCOURT
C 111	16.8	76.4	1086	13	CG954414	AGENCYCOURT
C 112	16.8	76.4	1092	13	CG954414	AGENCYCOURT
C 113	16.8	76.4	1098	13	CG954414	AGENCYCOURT
C 114	16.8	76.4	1104	13	CG954414	AGENCYCOURT
C 115	16.8	76.4	1110	13	CG954414	AGENCYCOURT
C 116	16.8	76.4	1116	13	CG954414	AGENCYCOURT
C 117	16.8	76.4	1122	13	CG954414	AGENCYCOURT
C 118	16.8	76.4	1128	13	CG954414	AGENCYCOURT
C 119	16.8	76.4	1134	13	CG954414	AGENCYCOURT
C 120	16.8	76.4	1140	13	CG954414	AGENCYCOURT
C 121	16.8	76.4	1146	13	CG954414	AGENCYCOURT
C 122	16.8	76.4	1152	13	CG954414	AGENCYCOURT
C 123	16.8	76.4	1158	13	CG954414	AGENCYCOURT
C 124	16.8	76.4	1164	13	CG954414	AGENCYCOURT
C 125	16.8	76.4	1170	13	CG954414	AGENCYCOURT
C 126	16.8	76.4	1176	13	CG954414	AGENCYCOURT
C 127	16.8	76.4	1182	13	CG954414	AGENCYCOURT
C 128	16.8	76.4	1188	13	CG954414	AGENCYCOURT
C 129	16.8	76.4	1194	13	CG954414	AGENCYCOURT
C 130	16.8	76.4	1200	13	CG954414	AGENCYCOURT
C 131	16.8	76.4	1206	13	CG954414	AGENCYCOURT
C 132	16.8	76.4	1212	13	CG954414	AGENCYCOURT
C 133	16.8	76.4	1218	13	CG954414	AGENCYCOURT
C 134	16.8	76.4	1224	13	CG954414	AGENCYCOURT
C 135	16.8	76.4	1230	13	CG954414	AGENCYCOURT
C 136	16.8	76.4	1236	13	CG954414	AGENCYCOURT
C 137	16.8	76.4	1242	13	CG954414	AGENCYCOURT
C 138	16.8	76.4	1248	13	CG954414	AGENCYCOURT
C 139	16.8	76.4	1254	13	CG954414	AGENCYCOURT
C 140	16.8	76.4	1260	13	CG954414	AGENCYCOURT
C 141	16.8	76.4	1266	13	CG954414	AGENCYCOURT
C 142	16.8	76.4	1272	13	CG954414	AGENCYCOURT
C 143	16.8	76.4	1278	13	CG954414	AGENCYCOURT
C 144	16.8	76.4	1284	13	CG954414	AGENCYCOURT
C 145	16.8	76.4	1290	13	CG954414	AGENCYCOURT
C 146	16.8	76.4	1296	13	CG954414	AGENCYCOURT
C 147	16.8	76.4	1302	13	CG954414	AGENCYCOURT
C 148	16.8	76.4	1308	13	CG954414	AGENCYCOURT
C 149	16.8	76.4	1314	13	CG954414	AGENCYCOURT
C 150	16.8	76.4	1320	13	CG954414	AGENCYCOURT

Db 172 CGAGCCAAACCTCGCCTAAC 151

RESULT 4
BJ185747/c
LOCUS
DEFINITION
BJ185747 390 bp mRNA linear EST 16-OCT-2003
caulonemata and malformed buds Physcomitrella patens subsp. patens
cDNA clone pphb7o24 5', mRNA sequence.
BJ185747
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Physcomitrella patens subsp. patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
1 (bases 1 to 390)
Nishiyama,T., Fujita,T., Shin-i,T., Seki,M., Nishide,H.,
Uchiyama,I., Kamiya,A., Carninci,P., Hayashizaki,Y., Shinozaki,K.,
Kohara,Y. and Hasebe,M.
Comparative genomics of Physcomitrella patens gametophytic
transcriptome and Arabidopsis thaliana: implication for land plant
evolution
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
22709184
JOURNAL MEDLINE
PUBMED
COMMENT
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tschini@genes.nig.ac.jp
A backbone of the vector is basically from pBluescript II (KS),
that was in vivo excised from a l-PIC phage vector (Carninci et al.
2001). 5' end of the cDNA that was digested with XhoI was ligated
to SalI site of the vector and the 3' end including polyA tail was
ligated to BamHI site of the
vector(5'- gagAGAGAGAGATCCAAACCTGGAgAGTTTTTTTITVTN-3' was
used as a 1st 3' primer, and
5'-ggTTCGAGTCATCGTGTCGACAGCGATGACTGCAGAACCGNNNN-3' as 2nd
5'-hairpin primer, giving the following 5' boarder sequence,
AGGCCAATCGCGAGCTCGATTGCTCGAGACCG). cDNA instert could be
amplified with conventional T7 and T3 primers. This full-length
cDNA library was generated according to the method described in
Nishiyama et al. (2003).
Protonemata were blended by the POLYTRON, and then cultivated on
the BCDATG medium for 13- 14 days under the continuous light.
These clones are available from RIKEN Bio Resource Center
(http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database
of Physcomitrella EST clones is available at the PHYSCOBASE
(http://moss.nibb.ac.jp).

FEATURES
Location/Qualifiers
source
1..390
/organism="Physcomitrella patens subsp. patens"
/mol_type="mRNA"
/sub_species="patens"
/db_xref="taxon:145481"
/clone="pphb7o24"
/tissue_type="mixture of chloronemata, caulonemata and
malformed buds"
/clone_lib="normalized full length cDNA library,
chloronemata, caulonemata and malformed buds"

ORIGIN
Query Match 85.5%; Score 18.8; DB 12; Length 390;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGCGAAAACCTCGCCTAAC 22
|||||
Db 162 CGAGCCAAACCTCGCCTAAC 141

Db 172 CGAGCCAAACCTCGCCTAAC 151

BJ185747 390 bp mRNA linear EST 16-OCT-2003
caulonemata and malformed buds Physcomitrella patens subsp. patens
cDNA clone pphb7o24 5', mRNA sequence.
BJ185747
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Physcomitrella patens subsp. patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
1 (bases 1 to 390)
Nishiyama,T., Fujita,T., Shin-i,T., Seki,M., Nishide,H.,
Uchiyama,I., Kamiya,A., Carninci,P., Hayashizaki,Y., Shinozaki,K.,
Kohara,Y. and Hasebe,M.
Comparative genomics of Physcomitrella patens gametophytic
transcriptome and Arabidopsis thaliana: implication for land plant
evolution
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
22709184
JOURNAL MEDLINE
PUBMED
COMMENT
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tschini@genes.nig.ac.jp
A backbone of the vector is basically from pBluescript II (KS),
that was in vivo excised from a l-PIC phage vector (Carninci et al.
2001). 5' end of the cDNA that was digested with XhoI was ligated
to SalI site of the vector and the 3' end including polyA tail was
ligated to BamHI site of the
vector(5'- gagAGAGAGAGATCCAAACCTGGAgAGTTTTTTTITVTN-3' was
used as a 1st 3' primer, and
5'-ggTTCGAGTCATCGTGTCGACAGCGATGACTGCAGAACCGNNNN-3' as 2nd
5'-hairpin primer, giving the following 5' boarder sequence,
AGGCCAATCGCGAGCTCGATTGCTCGAGACCG). cDNA instert could be
amplified with conventional T7 and T3 primers. This full-length
cDNA library was generated according to the method described in
Nishiyama et al. (2003).
Protonemata were blended by the POLYTRON, and then cultivated on
the BCDATG medium for 13- 14 days under the continuous light.
These clones are available from RIKEN Bio Resource Center
(http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database
of Physcomitrella EST clones is available at the PHYSCOBASE
(http://moss.nibb.ac.jp).

FEATURES
Location/Qualifiers
Source
1..390
/organism="Physcomitrella patens subsp. patens"
/mol_type="mRNA"
/sub_species="patens"
/db_xref="taxon:145481"
/clone="pphb7o24"
/tissue_type="mixture of chloronemata, caulonemata and
malformed buds"
/clone_lib="normalized full length cDNA library,
chloronemata, caulonemata and malformed buds"

ORIGIN
Query Match 85.5%; Score 18.8; DB 12; Length 390;
Best Local Similarity 90.9%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGCGAAAACCTCGCCTAAC 22
|||||
Db 162 CGAGCCAAACCTCGCCTAAC 141

RESULT 6

BJ175801/c
 LOCUS BJ175801 406 bp mRNA linear EST 16-OCT-2003
 DEFINITION BJ175801 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens cDNA clone pphb15h22 5', mRNA sequence.
 ACCESSION BJ175801
 VERSION BJ175801.1 GI:18343763
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 406)
 AUTHORS Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M.
 TITLE Comparative Genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution.
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
 MEDLINE 22709184
 PUBMED 12808149
 COMMENT Contact: Tadaasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'-GAGAGAGAGAGATCCAAACCTGGAGAGTTTTTTTTTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTGAGTCATCGTGTTCAGACGATGACTCGAGAACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGGCAATCGCCGAGCTCGAATGCTCGAGAACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).
 Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (<http://www.brc.riken.go.jp/lab/epd/eng/index.html>). The database of Physcomitrella EST clones is available at the PHYSCOBASE (<http://moss.nibb.ac.jp/>).
 FEATURES
 source
 1. 406
 Location/Qualifiers
 /organism="Physcomitrella patens subsp. patens"
 /mol_type="mRNA"
 /sub_species="patens"
 /db_xref="taxon:145481"
 /clone="pphb15h22"
 /tissue_type="mixture of chloronemata, caulonemata and malformed buds"
 /clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"
 ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 406;
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CCGGGCAAAACCTCGCCTAAC 22
 |||||
 DB 142 CCGAGCCAAACCTCGCCTAAC 121
 |||||
 RESULT 7
 BJ181492/c
 LOCUS BJ181492 420 bp mRNA linear EST 16-OCT-2003
 DEFINITION BJ181492 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens cDNA clone pphb30e14 5', mRNA sequence.
 ACCESSION BJ181492
 VERSION BJ181492.1 GI:18349443
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 420)
 AUTHORS Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M.
 TITLE Comparative Genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution.
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
 MEDLINE 22709184
 PUBMED 12808149
 COMMENT Contact: Tadaasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'-GAGAGAGAGAGATCCAAACCTGGAGAGTTTTTTTTTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTGAGTCATCGTGTTCAGACGATGACTCGAGAACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGGCAATCGCCGAGCTCGAATGCTCGAGAACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).
 Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (<http://www.brc.riken.go.jp/lab/epd/eng/index.html>). The database of Physcomitrella EST clones is available at the PHYSCOBASE (<http://moss.nibb.ac.jp/>).
 FEATURES
 source
 1. 420
 Location/Qualifiers
 /organism="Physcomitrella patens subsp. patens"
 /mol_type="mRNA"
 /sub_species="patens"
 /db_xref="taxon:145481"
 /clone="pphb30e14"
 /tissue_type="mixture of chloronemata, caulonemata and malformed buds"
 /clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"
 ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 420;
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CCGGGCAAAACCTCGCCTAAC 22
 |||||
 DB 162 CCGAGCCAAACCTCGCCTAAC 141
 |||||
 RESULT 8
 BJ183162/c
 LOCUS BJ183162 427 bp mRNA linear EST 16-OCT-2003
 DEFINITION BJ183162 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens cDNA clone pphb30e14 5', mRNA sequence.
 ACCESSION BJ183162
 VERSION BJ183162.1 GI:18349443
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 427)
 AUTHORS Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M.
 TITLE Comparative Genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution.
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
 MEDLINE 22709184
 PUBMED 12808149
 COMMENT Contact: Tadaasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'-GAGAGAGAGAGATCCAAACCTGGAGAGTTTTTTTTTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTGAGTCATCGTGTTCAGACGATGACTCGAGAACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGGCAATCGCCGAGCTCGAATGCTCGAGAACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).
 Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (<http://www.brc.riken.go.jp/lab/epd/eng/index.html>). The database of Physcomitrella EST clones is available at the PHYSCOBASE (<http://moss.nibb.ac.jp/>).
 FEATURES
 source
 1. 427
 Location/Qualifiers
 /organism="Physcomitrella patens subsp. patens"
 /mol_type="mRNA"
 /sub_species="patens"
 /db_xref="taxon:145481"
 /clone="pphb30e14"
 /tissue_type="mixture of chloronemata, caulonemata and malformed buds"
 /clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"
 ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 427;
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CCGGGCAAAACCTCGCCTAAC 22
 |||||
 DB 162 CCGAGCCAAACCTCGCCTAAC 141
 |||||

BJ181492 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens cDNA clone pphb30e14 5', mRNA sequence.
 ACCESSION BJ181492
 VERSION BJ181492.1 GI:18349443
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 420)
 AUTHORS Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M.
 TITLE Comparative Genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution.
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
 MEDLINE 22709184
 PUBMED 12808149
 COMMENT Contact: Tadaasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'-GAGAGAGAGAGATCCAAACCTGGAGAGTTTTTTTTTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTGAGTCATCGTGTTCAGACGATGACTCGAGAACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGGCAATCGCCGAGCTCGAATGCTCGAGAACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).
 Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (<http://www.brc.riken.go.jp/lab/epd/eng/index.html>). The database of Physcomitrella EST clones is available at the PHYSCOBASE (<http://moss.nibb.ac.jp/>).
 FEATURES
 source
 1. 420
 Location/Qualifiers
 /organism="Physcomitrella patens subsp. patens"
 /mol_type="mRNA"
 /sub_species="patens"
 /db_xref="taxon:145481"
 /clone="pphb30e14"
 /tissue_type="mixture of chloronemata, caulonemata and malformed buds"
 /clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"
 ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 420;
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CCGGGCAAAACCTCGCCTAAC 22
 |||||
 DB 162 CCGAGCCAAACCTCGCCTAAC 141
 |||||
 RESULT 8
 BJ183162/c
 LOCUS BJ183162 427 bp mRNA linear EST 16-OCT-2003
 DEFINITION BJ183162 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens cDNA clone pphb30e14 5', mRNA sequence.
 ACCESSION BJ183162
 VERSION BJ183162.1 GI:18349443
 KEYWORDS EST.
 SOURCE Physcomitrella patens subsp. patens
 ORGANISM Physcomitrella patens subsp. patens
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
 REFERENCE 1 (bases 1 to 427)
 AUTHORS Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M.
 TITLE Comparative Genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution.
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
 MEDLINE 22709184
 PUBMED 12808149
 COMMENT Contact: Tadaasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp
 A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phage vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'-GAGAGAGAGAGATCCAAACCTGGAGAGTTTTTTTTTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTGAGTCATCGTGTTCAGACGATGACTCGAGAACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGGCAATCGCCGAGCTCGAATGCTCGAGAACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).
 Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (<http://www.brc.riken.go.jp/lab/epd/eng/index.html>). The database of Physcomitrella EST clones is available at the PHYSCOBASE (<http://moss.nibb.ac.jp/>).
 FEATURES
 source
 1. 427
 Location/Qualifiers
 /organism="Physcomitrella patens subsp. patens"
 /mol_type="mRNA"
 /sub_species="patens"
 /db_xref="taxon:145481"
 /clone="pphb30e14"
 /tissue_type="mixture of chloronemata, caulonemata and malformed buds"
 /clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"
 ORIGIN
 Query Match 85.5%; Score 18.8; DB 12; Length 427;
 Best Local Similarity 90.9%; Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 CCGGGCAAAACCTCGCCTAAC 22
 |||||
 DB 162 CCGAGCCAAACCTCGCCTAAC 141
 |||||

CDNA clone pphb36107 5', mRNA sequence.

ACCESSION BJ183162

VERSION BJ183162.1 GI:18351111

KEYWORDS EST.

SOURCE Physcomitrella patens subsp. patens

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariaceae; Funariales; Funariaceae; Physcomitrella. 1 (bases 1 to 427)

REFERENCE Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Kohara, Y., and Hasebe, M. Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution

AUTHORS Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)

TITLE MEDLINE 22709184

JOURNAL PUBMED 12808149

COMMENT Contact: Tadasu Shin-i National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp

A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phase vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'-GAGAGAGAGAGATCCAACTGGAGTGTGTGTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTCGATGCTGTTCAGACGATGCTGAGAACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGGCAATCGCGAGCTGATTCGTCGAGAACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).

Protonemata were blended by the POLYTRON, and then cultivated on the BODATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (<http://www.brc.riken.go.jp/lab/epd/Eng/index.html>). The database of Physcomitrella EST clones is available at the PHYSCOBASE (<http://moss.nibb.ac.jp/>).

FEATURES

source

1. 427

Location/Qualifiers

/organism="Physcomitrella patens subsp. patens"

/mol_type="mRNA"

/sub_species="patens"

/db_xref="taxon:145481"

/clones="pphb36107"

/tissue_type="mixture of chloronemata, caulonemata and malformed buds"

/clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"

ORIGIN

Query Match 85.5%; Score 18.9; DB 12; Length 427;

Best Local Similarity 90.9%; Pred. No. 4.5e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGCAAAACCTCGCCTAAC 22

Db 166 CCGAGCCAAACCTCGCCTAAC 145

RESULT 9

BJ180094/c

LOCUS

DEFINITION

503 bp mRNA linear EST 16-OCT-2003

BJ180094 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens

CDNA clone pphb27c19 5', mRNA sequence.

ACCESSION BJ180094

CDNA clone pphb36107 5', mRNA sequence.

ACCESSION BJ180094

VERSION BJ180094.1 GI:18348047

KEYWORDS EST.

SOURCE Physcomitrella patens subsp. patens

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariaceae; Funariales; Funariaceae; Physcomitrella. 1 (bases 1 to 503)

REFERENCE Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H., Kohara, Y., and Hasebe, M. Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution

AUTHORS Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)

TITLE MEDLINE 22709184

JOURNAL PUBMED 12808149

COMMENT Contact: Tadasu Shin-i National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp

A backbone of the vector is basically from pBluescript II (KS), that was in vivo excised from a 1-FLC phase vector (Carninci et al. 2001). 5' end of the cDNA that was digested with XhoI was ligated to SalI site of the vector and the 3' end including polyA tail was ligated to BamHI site of the vector (5'-GAGAGAGAGAGATCCAACTGGAGTGTGTGTTTNN-3' was used as a 1st 3' primer, and 5'-GGTCTCGATGCTGTTCAGACGATGCTGAGAACCGNNNN-3' as 2nd 5'-hairpin primer, giving the following 5' boarder sequence, AGGCAATCGCGAGCTGATTCGTCGAGAACCG). cDNA insert could be amplified with conventional T7 and T3 primers. This full-length cDNA library was generated according to the method described in Nishiyama et al. (2003).

Protonemata were blended by the POLYTRON, and then cultivated on the BODATG medium for 13-14 days under the continuous light. These clones are available from RIKEN Bio Resource Center (<http://www.brc.riken.go.jp/lab/epd/Eng/index.html>). The database of Physcomitrella EST clones is available at the PHYSCOBASE (<http://moss.nibb.ac.jp/>).

FEATURES

source

1. 503

Location/Qualifiers

/organism="Physcomitrella patens subsp. patens"

/mol_type="mRNA"

/sub_species="patens"

/db_xref="taxon:145481"

/clones="pphb27c19"

/tissue_type="mixture of chloronemata, caulonemata and malformed buds"

/clone_lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"

ORIGIN

Query Match 85.5%; Score 18.8; DB 12; Length 503;

Best Local Similarity 90.9%; Pred. No. 4.6e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGCAAAACCTCGCCTAAC 22

Db 166 CCGAGCCAAACCTCGCCTAAC 145

RESULT 10

BJ198230/c

LOCUS

DEFINITION

564 bp mRNA linear EST 21-OCT-2003

BJ198230 normalized full length cDNA library, chloronemata, caulonemata and rhizoid-like protonemata Physcomitrella patens subsp. patens

CDNA clone pphn31m01 5', mRNA sequence.

ACCESSION BJ198230

VERSION

KEYWORDS

SOURCE Physcomitrella patens subsp. patens
ORGANISM Physcomitrella patens subsp. patens

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
AUTHORS Bryopsida; Funariidae; Funariaceae; Funariaceae; Physcomitrella.
1 (bases 1 to 564)

TITLE Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H.,
Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K.,
Kohara, Y. and Hasebe, M.
Comparative genomics of Physcomitrella patens gametophytic
transcriptome and Arabidopsis thaliana: implication for land plant
evolution

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
MEDLINE 22709184
PubMed 12808149

COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
A backbone of the vector is basically from pBluescript II (KS),
that was in vivo excised from a l-FLC phage vector (Carninci et al.
2001). 5' end of the cDNA that was digested with XhoI was ligated
to SalI site of the vector and the 3' end including polyA tail was
ligated to BamHI site of the
vector(5' - GAGAGAGAGAGATCCACCTGGAGAGTTTTTTTTTTTTTTVN-3' was
used as a 1st 3' primer, and
5'-ggttctcagtcacgctgttccagacagtcgactcgagacgcnnnn-3' as 2nd
5' hairpin primer, giving the following 5' boarder sequence,
AGGCCAATCGCGCGACTCGAATCTCGAGACCG). cDNA instert could be
amplified with conventional T7 and T3 primers. This full-length
cDNA library was generated according to the method described in
Nishiyama et al. (2003).
Protonemata were blended by the POLYTRON, and then cultivated on
the BCDATG medium for 13-14 days under the continuous light.
These clones are available from RIKEN Bio Resource Center.
(http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database
of Physcomitrella EST clones is available at the PHYSCOBASE
(http://moss.nibb.ac.jp).

FEATURES
source
1..564
Location/Qualifiers
/organism="Physcomitrella patens subsp. patens"
/mol_type="mRNA"
/sub_species="patens"
/db_xref="taxon:145481"
/clone="pphb37j17"
/tissue_type="mixture of chloronemata, caulonemata and
rhizoid-like protonemata"
/clone_libs="normalized full length cDNA library,
chloronemata, caulonemata and rhizoid-like protonemata"

ORIGIN
Query Match 85.5%; Score 19.8; DB 12; Length 564;
Best Local Similarity 90.9%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCGCTAAC 22
|||||
Db 174 CCGAGCCAAACCTCGCGCTAAC 153
|||||

RESULT 11
BU183579/c
LOCUS BU183579 586 bp mRNA linear EST 16-OCT-2003
DEFINITION BU183579 normalized full length cDNA library, chloronemata,
caulonemata and malformed buds Physcomitrella patens subsp. patens
cDNA clone pphb37j17 5', mRNA sequence.
ACCESSION BU183579
VERSION BU183579.1 GI:18351527
KEYWORDS EST.
SOURCE Physcomitrella patens subsp. patens
ORGANISM Physcomitrella patens subsp. patens

SOURCE Physcomitrella patens subsp. patens
ORGANISM Physcomitrella patens subsp. patens

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
AUTHORS Bryopsida; Funariidae; Funariaceae; Funariaceae; Physcomitrella.
1 (bases 1 to 564)

TITLE Nishiyama, T., Fujita, T., Shin-i, T., Seki, M., Nishide, H.,
Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K.,
Kohara, Y. and Hasebe, M.
Comparative genomics of Physcomitrella patens gametophytic
transcriptome and Arabidopsis thaliana: implication for land plant
evolution

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 8007-8012 (2003)
MEDLINE 22709184
PubMed 12808149

COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
A backbone of the vector is basically from pBluescript II (KS),
that was in vivo excised from a l-FLC phage vector (Carninci et al.
2001). 5' end of the cDNA that was digested with XhoI was ligated
to SalI site of the vector and the 3' end including polyA tail was
ligated to BamHI site of the
vector(5' - GAGAGAGAGAGATCCACCTGGAGAGTTTTTTTTTTTTTTVN-3' was
used as a 1st 3' primer, and
5'-ggttctcagtcacgctgttccagacagtcgactcgagacgcnnnn-3' as 2nd
5' hairpin primer, giving the following 5' boarder sequence,
AGGCCAATCGCGCGACTCGAATCTCGAGACCG). cDNA instert could be
amplified with conventional T7 and T3 primers. This full-length
cDNA library was generated according to the method described in
Nishiyama et al. (2003).
Protonemata were blended by the POLYTRON, and then cultivated on
the BCDATG medium for 13-14 days under the continuous light.
These clones are available from RIKEN Bio Resource Center.
(http://www.brc.riken.go.jp/lab/epd/Eng/index.html). The database
of Physcomitrella EST clones is available at the PHYSCOBASE
(http://moss.nibb.ac.jp).

FEATURES
source
1..564
Location/Qualifiers
/organism="Physcomitrella patens subsp. patens"
/mol_type="mRNA"
/sub_species="patens"
/db_xref="taxon:145481"
/clone="pphb37j17"
/tissue_type="mixture of chloronemata, caulonemata and
rhizoid-like protonemata"
/clone_libs="normalized full length cDNA library,
chloronemata, caulonemata and rhizoid-like protonemata"

ORIGIN
Query Match 85.5%; Score 19.8; DB 12; Length 564;
Best Local Similarity 90.9%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCGCTAAC 22
|||||
Db 174 CCGAGCCAAACCTCGCGCTAAC 153
|||||

RESULT 12
BH375652
LOCUS BH375652 666 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-17824.TR ND-TAM Anopheles gambiae genomic clone
AG-ND-17824, genomic survey sequence.
ACCESSION BH375652
VERSION BH375652.1 GI:17321794
KEYWORDS GSS.
SOURCE Anopheles gambiae (African malaria mosquito)
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.

REFERENCE 1 (bases 1 to 666)
 AUTHORS Hong,Y.S., Hogan,J.R., Wang,X., Sarkar,A., Sim,C., Loftus,B.J., Ren,C., Huff,E.R., Carlile,J.L., Black,K., Zhang,H.-B., Gardner,M.J. and Collins,F.H.
 TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito *Anopheles gambiae*
 JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)
 MEDLINE 22542063
 PUBMED 12655398
 COMMENT Other GSSs: AG-ND-178P24.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.
 Location/Qualifiers
 1..666
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-178P24"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"

FEATURES
 source
 1..666

ORIGIN
 Query Match 85.5%; Score 18.8; DB 28; Length 666;
 Best Local Similarity 90.9%; Pred. No. 4.8e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCCGCCTAAC 22
 ||||||||||||||||||||
 Db 521 CCGGGCAAAACCTCCGCCTAAC 542

RESULT 13
 BE226036
 LOCUS
 DEFINITION iai3a09.vi Mouse E10 5 12 5 Pancreas cDNA Library Mus musculus cDNA clone IMAGE:5638168 5', similar to SM.TPMZ_RAT P18344 TROPOMYOSIN ALPHA CHAIN, BRAIN-3 ;, mRNA sequence.

ACCESSION BE226036
 VERSION BE226036.1 GI:8931272
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 326)
 Melton,D., Meadows,A., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blisstein,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
 WashU-Harvard Pancreas EST Project
 Unpublished (2000)
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Adimika Meadows (melton@fas.harvard.edu)
 MGI:1847192 this sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 323.
 Location/Qualifiers
 1..326
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="ICR"
 /db_xref="taxon:10090"
 /clone="IMAGE:5638168"
 /sex="Both"
 /tissue_type="Pancreatic Bud"
 /dev_stage="Embryonic day 10.5 and 12.5, mixed"
 /lab_host="DH10B"
 /clone_lib="Mouse E10 5 12 5 Pancreas cDNA Library"
 /note="Organ: Pancreas; Vector: pSPORT1; Site 1: Not 1; Site 2: Sal 1; Library constructed using SuperScript plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.47 kb. Primary library, unamplified. cDNA Library Preparation: Guolin Chen."

FEATURES
 source
 1..326

ORIGIN
 Query Match 80.9%; Score 17.8; DB 10; Length 326;
 Best Local Similarity 90.5%; Pred. No. 1.2e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CGGGCAAAACCTCCGCCTAAC 22
 ||||||||||||||||||||
 Db 73 CGGGCAAAACCTCCGCCTAAC 93

RESULT 14
 BM944427
 LOCUS
 DEFINITION UI-M-EH0p-bvr-c-22-0-UI.r1 NIH_BMAP_EH0p Mus musculus cDNA clone IMAGE:5695821 5', mRNA sequence.

ACCESSION BM944427.1 GI:19428012
 VERSION BM944427
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 394)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
 source
 1..394
 /organism="Mus musculus"

```

/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5695821"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP EH0p"
/notes="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

```

ORIGIN

Query Match 80.9%; Score 17.8; DB 12; Length 394;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCTTAAC 22
|||||
DB 70 CGAGCAAAACCTCCGCTTAAC 90

RESULT 15
BF546805 399 bp mRNA linear EST 11-DEC-2000
LOCUS UI-R-C2p-qv-f-01-0-UI.r1 UI-R-C2p Rattus norvegicus cDNA clone
DEFINITION UI-R-C2p-qv-f-01-0-UI 5', mRNA sequence.
BF546805
VERSION BF546805.1 GI:11637912
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 399)
Bernaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 9704447
MEDLINE 8889548
PUBMED
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LLNL (info@image.llnl.gov). IMAGE ID= 1789368 The following
repetitive elements were found in this cDNA sequence: 54-145,
>BI-FHSINE/Alu
Seq primer: M13 Forward.
Location/Qualifiers
1. .399
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"

```

/db_xref="taxon:10116"
/clone="UI-R-C2p-qv-f-01-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-C2p"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bernaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

```

ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 399;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCTTAAC 22
|||||
DB 322 CGGGCAAAACCTCCGCTTAAC 342

RESULT 16
BE226096 412 bp mRNA linear EST 13-MAR-2002
LOCUS Ia14a08.Y1 Mouse E10 5 12 5 Pancreas cDNA library Mus musculus cDNA
DEFINITION clone IMAGE:5638262 5', similar to SW.TPMZ_RAT P18344 TROPOMYOSIN
ALPHA CHAIN, BRAIN-3 ;, mRNA sequence.
BE226096
VERSION BE226096.1 GI:8931332
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 412)
Wyllie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.
WashU-Harvard Pancreas EST Project
Unpublished (2000)
JOURNAL
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Adimika Meadows
(meadows@fas.harvard.edu)
MGI:1847286 this sequence now available from the IMAGE consortium,

for clone orders contact: info@image.llnl.gov
 Seq primer: -40RP from Gibco
 High quality sequence stop: 403.
 Location/Qualifiers
 1. .412

FEATURES

source
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="ICR"
 /db_xref="taxon:10090"
 /clone="IMAGE:5638262"
 /sex="Both"
 /tissue_type="Pancreatic Bud"
 /dev_stage="Embryonic day 10.5 and 12.5, mixed"
 /lab_host="DHI08"
 /clone_lib="Mouse E10 5 12 5 Pancreas cDNA Library"
 /note="Organ: Pancreas; Vector: pSPOR1; Site 1: Not 1; Site 2: Sal 1; Library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.47 kb. Primary library, unamplified. cDNA Library Preparation: Guolin Chen."
 ORIGIN
 Query Match 80.9%; Score 17.8; DB 10; Length 412;
 Best Local Similarity 90.5%; Pred. No. 1.3e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 CGGCAAAACCTCGCCTAAC 22
 |||||
 Db 73 CGAGCAAAACCGCGCCTAAC 93
 RESULT 17
 AI465361/c
 LOCUS
 DEFINITION
 mu75903.x1 Stratagene mouse embryonic carcinoma (#937317) Mus
 musculus cDNA clone IMAGE:651412 3' similar to SW:TPMZ_RAT P18344
 TROPOMYOSIN ALPHA CHAIN, BRAIN-3 ; mRNA sequence.
 ACCESSION
 AI465361 GI:4319391
 VERSION
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 441)
 AUTHORS
 Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
 Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
 Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
 Waterston,R. and Wilson,R.
 TITLE
 The WashU-NCI Mouse EST Project 1999
 JOURNAL
 Unpublished (1999)
 COMMENT
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 This clone was previously sequenced on the 5' end only, this new
 data is from the 3' end
 High quality sequence stop: 396.
 Location/Qualifiers
 1. .441
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:651412"
 /tissue_type="carcinoma"
 /dev_stage="embryonic"
 /lab_host="SOLR (kanamycin resistant)"

/clone_lib="Stratagene mouse embryonic carcinoma
 (#937317)"
 /note="vector: pbluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; Cloned unidirectionally. Primer: Oligo dr. P19 cell
 line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
 adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 80.9%; Score 17.8; DB 9; Length 441;
 Best Local Similarity 90.5%; Pred. No. 1.3e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 CGGCAAAACCTCGCCTAAC 22
 |||||
 Db 352 CGAGCAAAACCGCGCCTAAC 332
 RESULT 18
 AI265329/c
 LOCUS
 DEFINITION
 uk01908.y1 Schiller mouse MAC13 Mus musculus cDNA clone
 IMAGE:1958174 5' similar to GB:M19267 TROPOMYOSIN, FIBROBLAST
 ISOFORM TM3 (HUMAN); GB:X64831 M.musculus mRNA for skeletal muscle
 alpha tropomyosin (MOUSE); mRNA sequence.
 ACCESSION
 AI265329
 VERSION
 EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 458)
 AUTHORS
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Fan,P., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE
 The WashU-HMI Mouse EST Project
 JOURNAL
 Unpublished (1996)
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 431.
 Location/Qualifiers
 1. .458
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:1958174"
 /tissue_type="colon cancer"
 /cell_line="colon cancer cell line MAC13"
 /lab_host="SOLR"
 /clone_lib="Schiller mouse MAC13"
 /note="vector: pBluescript SK- (Stratagene); Site 1:
 EcoRI; Site 2: XhoI; Double-stranded cDNA was prepared
 from cell line MAC13 using primer
 5'-GAGAGAGAGAGAGAGAGAACTAGTCTGAGT(18)-3'. An EcoRI
 adaptor was used on the 5' end of the cDNA as follows:
 5'-AATTCGGCAGAG-3'. The library was size-selected and
 went through one round of amplification. Average insert
 size is 1.7 kb, with a range from 0.4-12 kb. This library
 was constructed by Dr. Martin Schiller (Johns Hopkins
 University)."

ORIGIN

Query Match 80.9%; Score 17.8; DB 9; Length 458;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCCTAAC 22
Db 221 CGAGCAAAACCGCGCCTAAC 201

RESULT 19
BE963645
LOCUS
DEFINITION BE863645 467 bp mRNA linear EST 29-SEP-2000
UI-M-BHO-ake-e-10-0-UI-r1 NIH_EMAP_M.S1 Mus musculus cDNA clone
UI-M-BHO-ake-e-10-0-UI 5', mRNA sequence.
ACCESSION BE863645
VERSION BE863645
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 467)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PubMed 8889548
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-1790, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov

CDNA Library Preparation: M.B. Soares lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements

FEATURES

source
Location/Qualifiers
1..467
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BHO-ake-e-10-0-UI"
/dev_stages="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH BMAP M.S1"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_EMAP_M.S1 library is a subtracted library derived from
a mixture of normalized libraries from ten regions of the
mouse brain (cerebellum, brain stems, olfactory bulb,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus). The driver used for
subtraction consisted of a pool of 20,000 cDNA clones
obtained from non-normalized and normalized libraries of
these ten regions of the mouse brain."

ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 467;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCCTAAC 22
Db 83 CGAGCAAAACCGCGCCTAAC 103

RESULT 20
AA584005

LOCUS

DEFINITION

AA584005

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Stratagene, Inc.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert length: 1079 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 147.

Location/Qualifiers

1..484

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1088730"

/tissue_type="larynx"

/lab_host="SOLR (kanamycin resistant)"

/clone_lib="NCI CGAP Lari"

/note="Organ: larynx; Vector: Bluescript SK-; Site 1:

ECORI; Site 2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Larynx. 5' adaptor sequence: 5' GAATTCGGCAG

3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'

Average insert size: 0.9 kb."

ORIGIN

Query Match 80.9%; Score 17.8; DB 9; Length 484;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCCGCCTAA 21

Db 206 CCGGCAAAACCTCCGCCTAA 226

RESULT 21

LOCUS

DEFINITION

CD807506

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

CD807506 487 bp mRNA linear EST 15-JUL-2003
UI-M-GW0-cio-p-20-0-UI-r1 NIH_EMAP_GW0 Mus musculus cDNA clone
IMAGE:30542491 5', mRNA sequence.
CD807506
CD807506.1 GI:32466332
EST.
Mus musculus (house mouse)
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 487)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES

Seq primer: pyX-5.
 Location/Qualifiers
 1..487

/organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:30542491"
 /tissue_type="whole eye"
 /dev_stage="embryo 15.5,16.5,17.5,18.5 dpc"
 /lab_hosts="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_GW0"
 /note="Organ: Eye; Vector: pyX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pyX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CTGGTCTC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN

Query Match 80.9%; Score 17.8; DB 14; Length 487;
 Best Local Similarity 90.5%; Pred. No. 1.3e+03;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CCGGCAAAACCTCGCCTAC 22

Db 167 CGAGCAAAACCGCGCTAAC 187

RESULT 22

R48986/c
 LOCUS R48986 500 bp mRNA linear EST 22-MAY-1995
 DEFINITION Y95501.s1 Soares infant brain INIB Homo sapiens cDNA clone
 IMAGE:38159 3', mRNA sequence.

ACCESSION R48986

VERSION R48986.1 GI:817750

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 500)

Hallier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,

Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,

Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,

Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and

Wilson,R.

Seq primer: pyX-5.

Location/Qualifiers

1..510

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

Seq primer: pyX-5.

Location/Qualifiers

1..510

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

Email: est@watson.wustl.edu

Insert Size: 1350

High quality sequence stops: 326 Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1350 Std Error: 0.00

Seq primer: Promega -21m13

High quality sequence stop: 326.

FEATURES

Location/Qualifiers

1..500

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="GDB:410700"

/db_xref="taxon:9606"

/clone="IMAGE:38159"

/sex="female"

/dev_stage="73 days post natal"

/lab_host="DH10B (ampicillin resistant)"

/clone_lib="Soares infant brain INIB"

/note="Organ: whole brain; Vector: Lactid BA; Site 1: Not

I; Site 2: Hind III; 1st strand cDNA was primed with a Not

I - oligo(dT) primer [5'

RACTGGAAGAAATCGCGCGCAGGAATTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Hind III adaptors

(Pharmacia), digested with Not I and directionally cloned

into the Not I and Hind III sites of the Lactid BA vector.

Library went through one round of normalization. Library

constructed by Bento Soares and M.Patima Bonaldo."

ORIGIN

Query Match 80.9%; Score 17.8; DB 14; Length 500;

Best Local Similarity 90.5%; Pred. No. 1.3e+03;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGCAAAACCTCGCCTAA 21

Db 395 CCGGCAAAACCTCGCCTGA 375

RESULT 23

BM950311

LOCUS BM950311

DEFINITION

IMAGE:5686916 5', mRNA sequence.

ACCESSION BM950311

VERSION BM950311.1 GI:19433901

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 510)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-remail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pyX-5.

Location/Qualifiers

1..510

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

Seq primer: pyX-5.

Location/Qualifiers

1..510

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

```

/clone="IMAGE:5686916"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP EHO"
/notes="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCAGCAG. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      80.9%; Score 17.8; DB 12; Length 510;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCCTAAC 22
DB 187 CGAGCAAAACCGCGCCTAAC 207

RESULT 24
A1598373/c
LOCUS
DEFINITION
EMBL250076 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
REMDQ94 3' end, mRNA sequence.
A1598373
VERSION
A1598373.1 GI:4607421
KEYWORDS
EST.
SOURCE
Rattus sp.
ORGANISM
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 532)
Lee,N.H., Glodok,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
Gene Index
JOURNAL
Unpublished (1998)
CONTACT
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1..532
/organism="Rattus sp."
/mol_type="mRNA"
/db_xref="taxon:10118"
/clone="REMDQ94"
/dev_stage="embryo 8, 12, 18 dpc"
/clone_lib="Normalized rat embryo, Bento Soares"
/notes="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"

ORIGIN
Query Match      80.9%; Score 17.8; DB 9; Length 532;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCCTAAC 22

```

```

DB 403 CGGAAAAACCTCCCTTAAC 383

RESULT 25
BU059049
LOCUS
DEFINITION
UI-M-FRO-cal-k-23-0-J1.r1 NIH_BMAP_FRO Mus musculus cDNA clone
IMAGE:6414094 5', mRNA sequence.
BU059049
ACCESSION
BU059049.1 GI:22499338
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1..606
/organism="Mus musculus"
/mol_type="mRNA"
/strains="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6414094"
/tissue_type="whole brain"
/dev_stage="embryo 13.5, 14.5, 16.5, 17.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FRO"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      80.9%; Score 17.8; DB 13; Length 606;
Best Local Similarity 90.5%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCCTAAC 22
DB 283 CGAGCAAAACCGCGCCTAAC 303

RESULT 26
CK134467
LOCUS
DEFINITION
CK134467
RE4552.3prime RE Drosophila melanogaster, normalized Embryo pFic-1

```

Drosophila melanogaster cDNA clone RE64552 3, mRNA sequence.
CK134467
VERSION CK134467.1 GI:38626403
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 658)
AUTHORS Scapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guatin, H., Harris, N., Li, P., Liao, G., Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and Rubin, G.M.
TITLE BDGP/HMI RE Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Other ESTs: RE64552.5prime
Contact: Stapleton, M.
BDGP Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE003486: arm:X [11180449,11484096]
estimated-cyto:10811-10D4: 01/30/2002
Plate: RE.645 row: E column: 4
High quality sequence stop: 480.
Location/Qualifiers
1..658
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clones="RE64552"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DMS-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFic-1"
/note="Organ: embryo; Vector: pFic1; Site:1: XhoI; Site:2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."
ORIGIN
Query Match 80.9%; Score 17.8; DB 14; Length 658;
Best Local Similarity 90.5%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CCGGGCAAAACCTCCGCTAA 21
Db 504 CCGGGCAAAACCTCCGCTGA 524
RESULT 27
BZ253738 697 bp DNA linear GSS 12-OCT-2002
LOCUS BZ230-306P22.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
DEFINITION CH230-306P22, genomic survey sequence.
ACCESSION BZ253738
VERSION BZ253738.1 GI:23915075
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tseng, G., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment

Unpublished (1999)
Other GSSs: CH230-306P22.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACFAC Resources (http://www.chori.org/bacpac/orering_information.htm). BAC end page: http://www.tigr.org/tadb/bac_ends/rat/bac_end_intro.html
Plate: 306 row: P column: 22
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..697
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strains="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-306P22"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI; CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by Pieter de Jong"
ORIGIN
Query Match 80.9%; Score 17.8; DB 28; Length 697;
Best Local Similarity 90.5%; Pred. No. 1.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CCGGGCAAAACCTCCGCTAA 21
Db 468 CCGGGCAAAACATCCCTTAA 488
RESULT 28
BQ863377/c 714 bp mRNA linear EST 14-AUG-2002
LOCUS BQ863377 Y9.ab1 QG-ABCDI lettuce salinas Lactuca sativa cDNA clone
DEFINITION QG23LI2, mRNA sequence.
ACCESSION BQ863377
VERSION BQ863377.1 GI:22248842
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Cichorioideae; Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 714)
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolman, J., Slabaugh, W.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project
JOURNAL Unpublished (2002)
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig2235, see <http://cgdb.ucdavis.edu/> for details.

Search completed: June 20, 2004, 14:14:19
Job time : 1055.06 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 24.328 Seconds
(without alignments)
501.846 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22

Sequence: 1 ccgggcaaaacctcgccctaac 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A COMB.seq:*

2: /cgn2_6/prodata/2/ina/5B COMB.seq:*

3: /cgn2_6/prodata/2/ina/6A COMB.seq:*

4: /cgn2_6/prodata/2/ina/6B COMB.seq:*

5: /cgn2_6/prodata/2/ina/6C COMB.seq:*

6: /cgn2_6/prodata/2/ina/6D COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	100.0	4403765	3	US-09-103-840A-2
C 2	22	100.0	4411529	3	US-09-103-840A-1
C 3	15.8	71.8	954	3	US-08-680-506-5
C 4	15.8	71.8	1347	3	US-08-680-506-8
C 5	15.8	71.8	1587	3	US-08-680-506-6
C 6	15.8	71.8	2233	3	US-08-680-506-4
C 7	15.8	71.8	5027	3	US-08-680-506-2
C 8	15.6	70.9	876	4	US-09-328-352-1452
C 9	15.6	70.9	1041	4	US-09-266-965-38
C 10	15.6	70.9	1206	4	US-09-489-039A-3575
C 11	15.6	70.9	2044	6	5229279-3
C 12	15.6	70.9	2094	6	5512669-3
C 13	15.6	70.9	19513	4	US-10-204-708-40
C 14	15.6	70.9	31940	4	US-09-596-002-13
C 15	15.6	70.9	33500	4	US-09-366-965-76
C 16	15.2	69.1	273	4	US-09-489-039A-2975
C 17	15.2	69.1	447	4	US-09-489-039A-2954
C 18	15.2	69.1	534	4	US-09-107-532A-2574
C 19	15.2	69.1	1588	4	US-09-976-594-625
C 20	15.2	69.1	1791	4	US-09-489-039A-2299
C 21	15.2	69.1	1830121	4	US-09-557-884-1
C 22	15.2	69.1	1830121	4	US-09-643-990A-1
C 23	15	68.2	6642	1	US-08-727-034-5
C 24	15	68.2	6843	1	US-08-727-034-6
C 25	14.8	67.3	774	4	US-08-858-207A-177
C 26	14.8	67.3	1038	4	US-09-328-352-715
C 27	14.8	67.3	1206	4	US-09-489-039A-1114

C 28	14.8	67.3	1482	4	US-09-543-681A-445	Sequence 445, App
C 29	14.8	67.3	1660	4	US-09-620-312D-723	Sequence 723, App
C 30	14.8	67.3	1770	4	US-09-489-039A-1224	Sequence 1224, App
C 31	14.8	67.3	4597	4	US-09-221-017B-1093	Sequence 1093, App
C 32	14.8	67.3	5004	4	US-09-489-039A-4453	Sequence 4453, App
C 33	14.8	67.3	5467	4	US-09-976-594-735	Sequence 735, App
C 34	14.8	67.3	28171	4	US-08-961-527-22	Sequence 22, Appli
C 35	14.8	67.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 36	14.8	67.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 37	14.6	66.4	63	3	US-07-921-104D-32	Sequence 32, Appli
C 38	14.6	66.4	345	4	US-09-489-039A-1576	Sequence 1576, Ap
C 39	14.6	66.4	351	3	US-07-921-104D-71	Sequence 71, Appli
C 40	14.6	66.4	375	3	US-07-921-104D-30	Sequence 30, Appli
C 41	14.6	66.4	506	4	US-09-621-976-2772	Sequence 2772, Ap
C 42	14.6	66.4	1203	4	US-09-489-039A-3904	Sequence 3904, Ap
C 43	14.6	66.4	1254	4	US-09-489-039A-4840	Sequence 4840, Ap
C 44	14.6	66.4	1488	4	US-03-252-931A-9307	Sequence 9307, Ap
C 45	14.6	66.4	1538	4	US-09-976-594-52	Sequence 52, Appli
C 46	14.6	66.4	1557	4	US-09-489-039A-4776	Sequence 4776, Ap
C 47	14.6	66.4	1569	4	US-09-724-623-57	Sequence 57, Appli
C 48	14.6	66.4	1602	4	US-09-252-931A-9270	Sequence 9270, Ap
C 49	14.6	66.4	1990	4	US-09-576-594-53	Sequence 53, Appli
C 50	14.6	66.4	2061	4	US-09-543-681A-1304	Sequence 1304, Ap
C 51	14.6	66.4	2992	4	US-03-221-017B-819	Sequence 819, App
C 52	14.6	66.4	3207	1	US-08-162-081B-35	Sequence 35, Appli
C 53	14.6	66.4	3207	1	US-08-780-872-35	Sequence 35, Appli
C 54	14.6	66.4	3207	1	US-09-085-957-35	Sequence 35, Appli
C 55	14.6	66.4	3240	1	US-08-162-081B-34	Sequence 34, Appli
C 56	14.6	66.4	3240	2	US-08-780-872-34	Sequence 34, Appli
C 57	14.6	66.4	3240	2	US-08-085-957-34	Sequence 34, Appli
C 58	14.6	66.4	3412	1	US-08-162-081B-32	Sequence 32, Appli
C 59	14.6	66.4	3412	2	US-08-780-872-32	Sequence 32, Appli
C 60	14.6	66.4	3412	3	US-09-085-957-32	Sequence 32, Appli
C 61	14.6	66.4	9432	1	US-08-277-231A-1	Sequence 1, Appli
C 62	14.6	66.4	9432	2	US-08-473-750-4	Sequence 4, Appli
C 63	14.6	66.4	9432	2	US-08-477-326-4	Sequence 4, Appli
C 64	14.4	65.5	250	4	US-08-956-171E-4568	Sequence 4568, Ap
C 65	14.4	65.5	252	4	US-09-134-001C-2253	Sequence 2253, Ap
C 66	14.4	65.5	356	4	US-08-956-171E-4293	Sequence 4293, Ap
C 67	14.4	65.5	502	4	US-08-956-171E-3907	Sequence 3907, Ap
C 68	14.4	65.5	609	4	US-08-956-171E-3937	Sequence 3937, Ap
C 69	14.4	65.5	649	4	US-09-470-191-5	Sequence 5, Appli
C 70	14.4	65.5	770	3	US-08-998-416-63	Sequence 63, Appli
C 71	14.4	65.5	969	4	US-09-134-001C-2283	Sequence 2283, Ap
C 72	14.4	65.5	1290	4	US-08-956-171E-3194	Sequence 3194, Ap
C 73	14.4	65.5	2047	3	US-08-836-261A-1	Sequence 1, Appli
C 74	14.4	65.5	4121	4	US-09-638-524A-1	Sequence 1, Appli
C 75	14.4	65.5	5476	4	US-10-204-708-81	Sequence 81, Appli
C 76	14.4	65.5	5666	4	US-10-204-708-30	Sequence 30, Appli
C 77	14.4	65.5	6156	4	US-10-204-708-60	Sequence 60, Appli
C 78	14.4	65.5	6343	3	US-08-581-148C-30	Sequence 30, Appli
C 79	14.2	64.5	384	4	US-09-489-039A-6139	Sequence 6139, Ap
C 80	14.2	64.5	462	4	US-09-252-931A-13639	Sequence 13639, A
C 81	14.2	64.5	471	4	US-09-489-039A-289	Sequence 289, App
C 82	14.2	64.5	501	4	US-09-252-931A-9251	Sequence 9251, Ap
C 83	14.2	64.5	513	4	US-09-489-039A-2764	Sequence 2764, Ap
C 84	14.2	64.5	516	4	US-09-252-931A-6461	Sequence 6461, Ap
C 85	14.2	64.5	525	4	US-09-252-931A-8915	Sequence 8915, Ap
C 86	14.2	64.5	525	4	US-09-489-039A-335	Sequence 335, App
C 87	14.2	64.5	552	4	US-09-489-039A-4449	Sequence 4449, Ap
C 88	14.2	64.5	594	4	US-09-252-931A-13801	Sequence 13801, A
C 89	14.2	64.5	603	4	US-09-252-931A-2672	Sequence 2672, Ap
C 90	14.2	64.5	621	4	US-09-252-931A-6380	Sequence 6380, Ap
C 91	14.2	64.5	714	4	US-09-252-931A-8702	Sequence 8702, Ap
C 92	14.2	64.5	735	4	US-08-961-527-366	Sequence 366, App
C 93	14.2	64.5	786	4	US-09-252-931A-2865	Sequence 2865, Ap
C 94	14.2	64.5	823	3	US-08-998-416-493	Sequence 493, App
C 95	14.2	64.5	840	4	US-09-489-039A-6039	Sequence 6039, Ap
C 96	14.2	64.5	873	4	US-09-252-931A-7015	Sequence 7015, Ap
C 97	14.2	64.5	906	4	US-09-252-931A-8799	Sequence 8799, Ap
C 98	14.2	64.5	933	4	US-09-252-931A-6169	Sequence 6169, Ap
C 99	14.2	64.5	1068	4	US-09-711-164-426	Sequence 246, App
C 100	14.2	64.5	1104	4	US-09-134-001C-2720	Sequence 2720, Ap

101	14.2	64.5	1203	4	US-09-252-991A-8410	Sequence 8410, App	174	14	63.6	1746	4	US-09-252-991A-13451	Sequence 13451, A
102	14.2	64.5	1221	4	US-09-252-991A-6296	Sequence 6296, App	175	14	63.6	1830	4	US-09-252-991A-8322	Sequence 8322, App
103	14.2	64.5	1239	4	US-09-489-039A-5782	Sequence 5782, App	176	14	63.6	2207	3	US-08-956-322-3	Sequence 3, Appli
C 104	14.2	64.5	1347	4	US-09-107-532A-308	Sequence 308, App	177	14	63.6	2210	3	US-08-464-700-53	Sequence 53, Appli
C 105	14.2	64.5	1350	4	US-09-252-991A-6012	Sequence 6012, App	178	14	63.6	2427	2	US-08-956-322-1	Sequence 1, Appli
C 106	14.2	64.5	1371	4	US-09-489-039A-1161	Sequence 1161, App	179	14	63.6	2427	2	US-08-678-039A-39	Sequence 39, Appli
C 107	14.2	64.5	1494	4	US-09-252-991A-13441	Sequence 13441, A	180	14	63.6	2436	4	US-09-199-637A-92	Sequence 92, Appli
C 108	14.2	64.5	1523	4	US-09-673-395A-1342	Sequence 1342, App	181	14	63.6	2535	4	US-09-672-737A-10	Sequence 10, Appli
C 109	14.2	64.5	1533	4	US-09-252-991A-800	Sequence 800, App	182	14	63.6	2730	1	US-08-344-536-1	Sequence 1, Appli
C 110	14.2	64.5	1569	4	US-09-252-991A-3005	Sequence 3005, App	183	14	63.6	2730	3	US-08-920-562-1	Sequence 1, Appli
C 111	14.2	64.5	1680	2	US-08-759-581B-3	Sequence 3, Appli	184	14	63.6	2730	3	US-08-920-562-1	Sequence 1, Appli
C 112	14.2	64.5	1680	4	US-09-304-711-3	Sequence 3, Appli	185	14	63.6	3750	4	US-09-252-991A-13820	Sequence 13820, A
C 113	14.2	64.5	1680	4	US-09-173-281-3	Sequence 3, Appli	186	14	63.6	3750	3	US-09-110-116-2	Sequence 2, Appli
C 114	14.2	64.5	1701	4	US-09-252-991A-1950	Sequence 1950, App	187	14	63.6	3750	4	US-09-252-991A-13627	Sequence 13627, A
C 115	14.2	64.5	1752	4	US-09-543-681A-1651	Sequence 1651, App	188	14	63.6	4421	2	US-08-257-963B-9	Sequence 9, Appli
C 116	14.2	64.5	1761	4	US-09-489-039A-1779	Sequence 1779, App	189	14	63.6	4421	4	US-08-367-841A-9	Sequence 9, Appli
C 117	14.2	64.5	1770	4	US-09-107-532A-2432	Sequence 2432, App	190	14	63.6	4421	4	US-08-520-373D-6	Sequence 6, Appli
C 118	14.2	64.5	1788	4	US-09-252-991A-863	Sequence 863, App	191	14	63.6	4421	5	PCT-US95-07201-9	Sequence 9, Appli
C 119	14.2	64.5	1804	1	US-08-306-691B-40	Sequence 40, Appli	192	14	63.6	6909	4	US-09-199-637A-111	Sequence 111, App
C 120	14.2	64.5	1804	4	US-09-167-322-14	Sequence 14, Appli	193	14	63.6	8962	4	US-09-028-934-28	Sequence 28, Appli
C 121	14.2	64.5	1804	5	PCT-US93-06251-82	Sequence 82, Appli	194	14	63.6	8962	4	US-09-691-861A-3	Sequence 3, Appli
C 122	14.2	64.5	1845	4	US-09-489-039A-2790	Sequence 2790, App	195	14	63.6	24707	4	US-09-740-037-3	Sequence 3, Appli
C 123	14.2	64.5	1935	4	US-09-252-991A-8438	Sequence 8438, App	196	14	63.6	42335	4	US-09-199-637A-1	Sequence 1, Appli
C 124	14.2	64.5	1977	4	US-09-252-991A-8339	Sequence 8339, App	197	14	63.6	45819	4	US-08-453-702B-72	Sequence 72, Appli
C 125	14.2	64.5	2097	4	US-09-252-991A-885	Sequence 885, App	198	14	63.6	11282	4	US-09-754-250-3	Sequence 3, Appli
C 126	14.2	64.5	2098	4	US-09-489-847-20	Sequence 20, Appli	199	14	63.6	38	4	US-09-371-772B-12522	Sequence 12522, A
C 127	14.2	64.5	2169	4	US-09-489-847-20	Sequence 20, Appli	200	13.8	62.7	48	4	US-09-453-702B-153	Sequence 153, App
C 128	14.2	64.5	2230	3	US-08-448-194-7	Sequence 7, Appli	201	13.8	62.7	254	4	US-09-016-434-743	Sequence 743, App
C 129	14.2	64.5	2230	4	US-08-867-921-7	Sequence 7, Appli	202	13.8	62.7	295	4	US-09-313-294A-4988	Sequence 4988, App
C 130	14.2	64.5	2260	2	US-08-788-750-1	Sequence 1, Appli	203	13.8	62.7	321	4	US-09-313-294A-4988	Sequence 4988, App
C 131	14.2	64.5	2298	4	US-09-252-991A-13540	Sequence 13540, A	204	13.8	62.7	402	4	US-09-634-238-96	Sequence 96, Appli
C 132	14.2	64.5	2397	4	US-09-489-039A-2874	Sequence 2874, App	205	13.8	62.7	484	4	US-09-328-475C-183	Sequence 183, App
C 133	14.2	64.5	2412	4	US-09-252-991A-7043	Sequence 7043, App	206	13.8	62.7	526	3	US-09-328-111-361	Sequence 261, App
C 134	14.2	64.5	2412	4	US-09-252-991A-13908	Sequence 13908, A	207	13.8	62.7	635	3	US-09-328-111-56	Sequence 56, Appli
C 135	14.2	64.5	2472	4	US-09-252-991A-7084	Sequence 7084, App	208	13.8	62.7	714	3	US-08-971-090-2	Sequence 2, Appli
C 136	14.2	64.5	2916	4	US-09-976-594-210	Sequence 210, App	209	13.8	62.7	714	3	US-08-971-090-3	Sequence 3, Appli
C 137	14.2	64.5	2946	4	US-09-252-991A-2045	Sequence 2045, App	210	13.8	62.7	933	4	US-09-252-991A-8606	Sequence 8606, App
C 138	14.2	64.5	3286	2	US-08-867-921-7	Sequence 7, Appli	211	13.8	62.7	1089	4	US-09-252-991A-1162	Sequence 1162, App
C 139	14.2	64.5	3509	2	US-08-817-436A-1	Sequence 1, Appli	212	13.8	62.7	1125	4	US-08-107-532A-2373	Sequence 2373, App
C 140	14.2	64.5	5228	4	US-10-164-595-23	Sequence 23, Appli	213	13.8	62.7	1131	6	5180810-3	Patent No. 5180810
C 141	14.2	64.5	9278	1	US-08-243-542-9	Sequence 9, Appli	214	13.8	62.7	1146	4	US-09-252-991A-8577	Sequence 8577, App
C 142	14.2	64.5	9278	1	US-08-477-407-9	Sequence 9, Appli	215	13.8	62.7	1182	4	US-09-543-681A-817	Sequence 817, App
C 143	14.2	64.5	9278	1	US-08-484-355-9	Sequence 9, Appli	216	13.8	62.7	1199	4	US-08-630-915A-15	Sequence 15, Appli
C 144	14.2	64.5	10357	4	US-08-961-527-191	Sequence 191, App	217	13.8	62.7	1260	4	US-09-328-352-848	Sequence 848, App
C 145	14.2	64.5	18912	5	PCT-US96-03916-59	Sequence 59, App	218	13.8	62.7	1326	4	US-09-328-352-3628	Sequence 3628, App
C 146	14.2	64.5	36063	4	US-08-311-731A-140	Sequence 140, App	219	13.8	62.7	1364	4	US-09-620-322D-36	Sequence 36, Appli
C 147	14.2	64.5	42157	4	US-08-311-731A-126	Sequence 126, App	220	13.8	62.7	1400	1	US-09-252-991A-11683	Sequence 11683, A
C 148	14.2	64.5	42157	4	US-09-356-575E-22	Sequence 22, App	221	13.8	62.7	1400	1	US-08-051-142-1	Sequence 1, Appli
C 149	14.2	64.5	45	3	US-09-252-991A-13101	Sequence 13101, A	222	13.8	62.7	1446	4	US-09-489-039A-5151	Sequence 5151, App
C 150	14.2	64.5	315	4	US-09-073-009-133	Sequence 133, App	223	13.8	62.7	1506	4	US-09-252-991A-224	Sequence 224, App
C 151	14.2	64.5	387	4	US-09-073-009-133	Sequence 133, App	224	13.8	62.7	1536	4	US-09-252-991A-11555	Sequence 11555, A
C 152	14.2	64.5	387	4	US-09-073-009-133	Sequence 133, App	225	13.8	62.7	1536	1	US-07-813-584A-1	Sequence 1, Appli
C 153	14.2	64.5	405	4	US-09-252-991A-13883	Sequence 13883, A	226	13.8	62.7	1560	1	US-08-330-515-1	Sequence 1, Appli
C 154	14.2	64.5	414	4	US-09-252-991A-13427	Sequence 13427, App	227	13.8	62.7	1613	3	US-08-725-594A-32	Sequence 32, Appli
C 155	14.2	64.5	546	4	US-09-199-637A-98	Sequence 98, App	228	13.8	62.7	1613	3	US-08-937-993-32	Sequence 32, Appli
C 156	14.2	64.5	608	4	US-09-221-017B-456	Sequence 456, App	229	13.8	62.7	1674	4	US-09-252-991A-1253	Sequence 1253, App
C 157	14.2	64.5	615	4	US-09-489-039A-375	Sequence 375, App	230	13.8	62.7	1784	6	5180810-2	Patent No. 5180810
C 158	14.2	64.5	713	3	US-09-252-991A-13717	Sequence 13717, A	231	13.8	62.7	1923	4	US-09-503-814A-5	Sequence 5, Appli
C 159	14.2	64.5	798	4	US-09-252-991A-13717	Sequence 13717, A	232	13.8	62.7	1923	4	US-09-503-814A-7	Sequence 7, Appli
C 160	14.2	64.5	809	3	US-09-252-991A-13717	Sequence 13717, A	233	13.8	62.7	1923	4	US-09-503-814A-7	Sequence 7, Appli
C 161	14.2	64.5	810	3	US-09-252-991A-13717	Sequence 13717, A	234	13.8	62.7	1959	4	US-09-252-991A-7983	Sequence 7983, App
C 162	14.2	64.5	933	4	US-09-252-991A-8451	Sequence 8451, App	235	13.8	62.7	2010	4	US-09-252-991A-8169	Sequence 8169, App
C 163	14.2	64.5	966	4	US-09-489-039A-2178	Sequence 2178, App	236	13.8	62.7	2230	1	US-08-200-512-1	Sequence 1, Appli
C 164	14.2	64.5	1011	4	US-09-252-991A-16115	Sequence 16115, A	237	13.8	62.7	2436	4	US-09-540-236-492	Sequence 492, App
C 165	14.2	64.5	1089	4	US-09-252-991A-13652	Sequence 13652, A	238	13.8	62.7	2513	4	US-09-585-173B-3	Sequence 3, Appli
C 166	14.2	64.5	1110	4	US-09-252-991A-8388	Sequence 8388, App	239	13.8	62.7	2642	4	US-09-595-002-4	Sequence 4, Appli
C 167	14.2	64.5	1131	4	US-09-252-991A-8486	Sequence 8486, App	240	13.8	62.7	2689	1	US-08-465-795-2	Sequence 2, Appli
C 168	14.2	64.5	1233	4	US-09-543-681A-2246	Sequence 2246, App	241	13.8	62.7	2736	4	US-09-252-991A-8666	Sequence 8666, App
C 169	14.2	64.5	1401	1	US-07-843-949A-3	Sequence 3, Appli	242	13.8	62.7	3245	1	US-07-935-311A-3	Sequence 3, Appli
C 170	14.2	64.5	1401	2	US-08-218-978-3	Sequence 3, Appli	243	13.8	62.7	3245	1	PCT-US93-07996-3	Sequence 3, Appli
C 171	14.2	64.5	1477	4	US-09-976-594-794	Sequence 794, App	244	13.8	62.7	3245	1	PCT-US93-07996-3	Sequence 3, Appli
C 172	14.2	64.5	1503	4	US-09-252-991A-13684	Sequence 13684, A	245	13.8	62.7	5092	4	US-09-620-312D-153	Sequence 153, App
C 173	14.2	64.5	1644	3	US-08-948-564-5	Sequence 5, Appli	246	13.8	62.7	5494	4	US-09-437-566A-45	Sequence 45, Appli


```

247 13.8 62.7 5889 1 US-08-465-795-1 Sequence 1, Appli
C 248 13.8 62.7 6669 4 US-10-204-708-6 Sequence 6, Appli
C 249 13.8 62.7 7001 1 US-08-258-261B-1 Sequence 1, Appli
C 250 13.8 62.7 7001 1 US-08-456-837-1 Sequence 1, Appli
C 251 13.8 62.7 7001 1 US-08-457-342-1 Sequence 1, Appli
C 252 13.8 62.7 7001 1 US-08-457-646A-1 Sequence 1, Appli
C 253 13.8 62.7 7001 1 US-08-458-076A-1 Sequence 1, Appli
C 254 13.8 62.7 7001 1 US-08-457-338A-1 Sequence 1, Appli
C 255 13.8 62.7 7001 1 US-08-761-258-6 Sequence 6, Appli
C 256 13.8 62.7 7001 1 US-08-729-214-1 Sequence 1, Appli
C 257 13.8 62.7 7001 1 US-08-977-306-6 Sequence 6, Appli
C 258 13.8 62.7 7001 3 US-09-028-934-1 Sequence 14, Appli
C 259 13.8 62.7 12886 4 US-09-453-702B-14 Sequence 14, Appli
C 260 13.8 62.7 65792 4 US-09-596-002-31 Sequence 31, Appli
C 261 13.8 62.7 1230025 4 US-09-198-452A-1 Sequence 1, Appli
C 262 13.8 62.7 1664976 4 US-08-916-421B-1 Sequence 1, Appli
C 263 13.6 61.8 35 1 US-08-361-337-20 Sequence 20, Appli
C 264 13.6 61.8 372 4 US-09-266-965-65 Sequence 65, Appli
C 265 13.6 61.8 390 4 US-09-134-000C-495 Sequence 495, Appli
C 266 13.6 61.8 405 4 US-09-489-039A-5545 Sequence 5545, Appli
C 267 13.6 61.8 537 4 US-09-489-039A-2077 Sequence 2077, Appli
C 268 13.6 61.8 542 3 US-08-818-112-15 Sequence 15, Appli
C 269 13.6 61.8 542 4 US-08-818-111-15 Sequence 15, Appli
C 270 13.6 61.8 542 4 US-09-056-556-15 Sequence 15, Appli
C 271 13.6 61.8 542 4 US-09-072-596-15 Sequence 15, Appli
C 272 13.6 61.8 542 4 US-09-072-967-15 Sequence 15, Appli
C 273 13.6 61.8 600 4 US-09-669-751-245 Sequence 245, Appli
C 274 13.6 61.8 633 4 US-09-252-991A-6234 Sequence 6234, Appli
C 275 13.6 61.8 636 4 US-09-429-906B-3 Sequence 3, Appli
C 276 13.6 61.8 753 4 US-09-252-991A-16180 Sequence 16180, A
C 277 13.6 61.8 759 4 US-09-252-991A-16011 Sequence 16011, A
C 278 13.6 61.8 771 4 US-09-489-039A-4829 Sequence 4829, Appli
C 279 13.6 61.8 783 4 US-09-252-991A-14448 Sequence 14448, A
C 280 13.6 61.8 786 4 US-09-489-039A-3554 Sequence 3554, Appli
C 281 13.6 61.8 786 4 US-08-891-641-22 Sequence 22, Appli
C 282 13.6 61.8 939 4 US-09-252-991A-13394 Sequence 13394, A
C 283 13.6 61.8 960 4 US-09-328-352-3516 Sequence 3516, Appli
C 284 13.6 61.8 962 4 US-08-936-165A-140 Sequence 140, Appli
C 285 13.6 61.8 979 4 US-09-252-991A-13601 Sequence 13601, A
C 286 13.6 61.8 1008 4 US-09-489-039A-5751 Sequence 5751, Appli
C 287 13.6 61.8 1017 4 US-09-345-236B-45 Sequence 44, Appli
C 288 13.6 61.8 1017 4 US-09-345-236B-45 Sequence 46, Appli
C 289 13.6 61.8 1029 4 US-09-252-991A-6399 Sequence 6399, Appli
C 290 13.6 61.8 1030 3 US-08-858-003-2 Sequence 2, Appli
C 291 13.6 61.8 1030 3 US-09-078-166-2 Sequence 2, Appli
C 292 13.6 61.8 1030 3 US-08-397-467-2 Sequence 2, Appli
C 293 13.6 61.8 1038 4 US-09-543-681A-2620 Sequence 2620, Appli
C 294 13.6 61.8 1105 4 US-09-072-596-329 Sequence 329, Appli
C 295 13.6 61.8 1105 4 US-09-072-967-334 Sequence 334, Appli
C 296 13.6 61.8 1138 4 US-10-164-595-33 Sequence 33, Appli
C 297 13.6 61.8 1140 4 US-09-543-681A-3785 Sequence 3785, Appli
C 298 13.6 61.8 1154 4 US-09-252-991A-6068 Sequence 6068, Appli
C 299 13.6 61.8 1209 4 US-09-252-991A-137 Sequence 137, Appli
C 300 13.6 61.8 1239 4 US-09-543-681A-4010 Sequence 4010, Appli

```

ALIGNMENTS

```

RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

```

```

; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "a" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

```

Query Match 100.0%; Score 22; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CCGGCAAAACCTCGCCTAAC 22
DB 3234122 CCGGCAAAACCTCGCCTAAC 3234101

```

RESULT 2

```

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Query Match 100.0%; Score 22; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CCGGCAAAACCTCGCCTAAC 22
DB 3239846 CCGGCAAAACCTCGCCTAAC 3239825

```

RESULT 3

```

US-08-680-506-5
; Sequence 5, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: REYNOLDS, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-5

```

Query Match 71.8%; Score 15.8; DB 3; Length 954;
Best Local Similarity 89.5%; Pred. No. 76;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCGGCT 19
Db 227 CCGGGCAAAATCGCGGCT 245

RESULT 4

US-08-680-506-8
; Sequence 8, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-8

Query Match 71.8%; Score 15.8; DB 3; Length 1347;
Best Local Similarity 89.5%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCGGCT 19
Db 227 CCGGGCAAAATCGCGGCT 245

RESULT 5

US-08-680-506-6
; Sequence 6, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1387
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-6

Query Match 71.8%; Score 15.8; DB 3; Length 1587;
Best Local Similarity 89.5%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCGGCT 19
Db 307 CCGGGCAAAATCGCGGCT 325

RESULT 6

US-08-680-506-4
; Sequence 4, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.

; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2233
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-4

Query Match 71.8%; Score 15.8; DB 3; Length 2233;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCGGCT 19
Db 307 CCGGGCAAAATCGCGGCT 325

RESULT 7

US-08-680-506-2
; Sequence 2, Application US/08680506C
; Patent No. 6008013
; GENERAL INFORMATION:
; APPLICANT: Reynolds, Paul R.
; TITLE OF INVENTION: CHONDROCYTE PROTEINS
; FILE REFERENCE: 176/60091
; CURRENT APPLICATION NUMBER: US/08/680,506C
; CURRENT FILING DATE: 1996-07-08
; EARLIER APPLICATION NUMBER: 60/021,672
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5027
; TYPE: DNA
; ORGANISM: Gallus gallus
US-08-680-506-2

Query Match 71.8%; Score 15.8; DB 3; Length 5027;
Best Local Similarity 89.5%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CCGGGCAAAACCTCGGCT 19
Db 307 CCGGGCAAAATCGCGGCT 325

RESULT 8

US-09-328-352-1452/c
; Sequence 1452, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1452
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1452

Query Match 70.9%; Score 15.6; DB 4; Length 876;
Best Local Similarity 81.8%; Pred. No. 95;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

5229279-3/c
; Patent No. 5229279
; APPLICANT: PEOPLES, OLIVER P.;SINSKEY,ANTHONY J.
; TITLE OF INVENTION: METHOD FOR PRODUCING NOVEL POLYESTER
; BIOPOLYMERS
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/556,535
; FILING DATE: 13-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 67,695
; FILING DATE: 29-AUG-1987
; SEQ ID NO:3:
; LENGTH: 2044
5229279-3
Query Match 70.9%; Score 15.6; DB 6; Length 2044;
Best Local Similarity 81.8%; Pred. No. 1.1e-02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCCGCTAAC 22
Db 1263 CCGGCAGAACCGCGCTCAC 1242
RESULT 12
5512669-3/c
; Patent No. 5512669
; APPLICANT: PEOPLES, OLIVER P.;SINSKEY, ANTHONY J.
; TITLE OF INVENTION: GENE ENCODING BACTERIAL ACETOACETYL-COA
; REDUCTASE
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,667
; FILING DATE: 29-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 124,570
; FILING DATE: 20-SEP-1993
; APPLICATION NUMBER: 944,488
; FILING DATE: 03-NOV-1992
; APPLICATION NUMBER: 566,535
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 67,695
; FILING DATE: 29-JUN-1987
; SEQ ID NO:3:
; LENGTH: 2094
5512669-3
Query Match 70.9%; Score 15.6; DB 6; Length 2094;
Best Local Similarity 81.8%; Pred. No. 1.1e-02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCCGCTAAC 22
Db 1313 CCGGCAGAACCGCGCTCAC 1292
RESULT 13
US-10-204-708-40/c
; Sequence 40, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8

US-09-266-965-38
; Sequence 38, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-38
Query Match 70.9%; Score 15.6; DB 4; Length 1041;
Best Local Similarity 81.8%; Pred. No. 97;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCCGCTAAC 22
Db 506 CCGGGAACACATCCGCTCAC 527
RESULT 10
US-09-489-039A-3575
; Sequence 3575, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3575
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3575
Query Match 70.9%; Score 15.6; DB 4; Length 1206;
Best Local Similarity 81.8%; Pred. No. 99;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGCAAAACCTCCGCTAAC 22
Db 631 CTGGCAATATCTCCGCTATC 652
RESULT 11

; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 40
; LENGTH: 19513
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-40

Query Match 70.9%; Score 15.6; DB 4; Length 19513;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCCTAAC 22
DB 40 CCGATCAAAACCTCCACCAAC 19

RESULT 14
US-09-596-002-13/c
; Sequence 13, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 31940
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 13
; PUBLICATION INFORMATION:
US-09-596-002-13

Query Match 70.9%; Score 15.6; DB 4; Length 31940;
Best Local Similarity 81.8%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCCTAAC 22
DB 30387 CCGGGCAAAACCTCCGCAAC 30366

RESULT 15
US-09-266-965-76/c
; Sequence 76, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965

; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 76
; LENGTH: 53500
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-76

Query Match 70.9%; Score 15.6; DB 4; Length 53500;
Best Local Similarity 81.8%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGGCAAAACCTCGCCTAAC 22
DB 9734 CCGGGCAACACATCCGCTCAC 9713

RESULT 16
US-09-489-039A-2975
; Sequence 2975, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2975
; LENGTH: 273
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2975

Query Match 69.1%; Score 15.2; DB 4; Length 273;
Best Local Similarity 85.0%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 CCGGGCAAAACCTCGCCTAA 21
DB 205 CCGGGCAAAACCTCTCCGAA 224

RESULT 17
US-09-489-039A-2954/c
; Sequence 2954, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2954
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2954

Query Match 69.1%; Score 15.2; DB 4; Length 447;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CGGCAAAACCTCGCCTAA 21
Db 421 CCGCAAAACCTCGCGAA 402

RESULT 18
US-09-107-532A-2574/c
; Sequence 2574, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107.532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Denise
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 2574:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 534 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (b) LOCATION 1...534
; SEQUENCE DESCRIPTION: SEQ ID NO: 2574:
US-09-107-532A-2574

Query Match 69.1%; Score 15.2; DB 4; Length 534;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GGGCAAAACCTCGGCTAAC 22
Db 247 GGGCAAAACCGACTCTTAC 228

RESULT 19
US-09-976-594-625/c

; Sequence 625, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 625
; LENGTH: 1588
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1344185CBI
; LOCATION: 7, 19, 25, 32, 45, 59, 88, 96, 108, 116, 124, 150, 167, 181, 191, 216, 216, 359,
; LOCATION: 221, 226, 233, 246, 264, 276, 286, 303, 307, 315, 330, 332, 336, 351, 359,
; LOCATION: 363, 373, 377, 382, 395, 407, 424, 431, 435, 442, 446, 470, 488, 492, 499,
; LOCATION: 504, 537, 551, 556, 583, 592, 596, 604, 615, 641, 671, 696, 709, 714, 726,
; LOCATION: 738, 747, 751, 756, 782, 774, 810, 842, 844, 906
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-625

Query Match 69.1%; Score 15.2; DB 4; Length 1588;
Best Local Similarity 81.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGCAAAACCTCGCCTAA 21
Db 584 CNGGGCAAAACCCGCCCAAA 564

RESULT 20
US-09-489-039A-2299
; Sequence 2299, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2299
; LENGTH: 1791
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2299

Query Match 69.1%; Score 15.2; DB 4; Length 1791;
Best Local Similarity 85.0%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CGGCAAAACCTCGCCTAA 21
Db 1700 CGGTAAACCTCGCGAA 1719

RESULT 21
US-09-557-884-1/c
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 69.1%; Score 15.2; DB 4; Length 1830121;
Best Local Similarity 85.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CGGGCAAAACCTCCGCTAA 21
Db 430196 CGAGCAAAACCTTGGCTAA 430177

RESULT 22

US-09-643-990A-1/c
Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 69.1%; Score 15.2; DB 4; Length 1830121;
Best Local Similarity 85.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CGGGCAAAACCTCCGCTAA 21
Db 430196 CGAGCAAAACCTTGGCTAA 430177

RESULT 23

US-08-727-034-5
Sequence 5, Application US/08727034
Patent No. 5665872

GENERAL INFORMATION:
APPLICANT: SAITO, YASHUSHI
APPLICANT: IWASAKI, AKIO
APPLICANT: ARAI, KOICHI

APPLICANT: YAMAZAKI, HIROYUKI
TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND
TITLE OF INVENTION: THE GENE CODING THEREFOR

NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727,034
FILING DATE: 08-OCT-1996
CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 261440/1995
FILING DATE: 09-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 102451/1996
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 80-079-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 6642 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-727-034-5

Query Match 68.2%; Score 15; DB 1; Length 6642;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGGCGAAAACCTCCG 16
|||||
Db 5828 CGGCGAAAACCTCCG 5842

RESULT 24
US-08-727-034-6
Sequence 6, Application US/08727034
Patent No. 5665872
GENERAL INFORMATION:
APPLICANT: SAITO, YAGUSHI
APPLICANT: IWASAKI, AKIO
APPLICANT: ARAI, KOICHI
APPLICANT: YAMAZAKI, HIROYUKI
TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND
TITLE OF INVENTION: THE GENE CODING THEREFOR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727,034
FILING DATE: 08-OCT-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 261440/1995
FILING DATE: 09-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 102451/1996
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 80-079-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 6843 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS

LOCATION: 81..6725
OTHER INFORMATION: /note= "Identification Method: S"
FEATURE:
NAME/KEY: sig peptide
LOCATION: 81..164
OTHER INFORMATION: /note= "Identification Method: S"
FEATURE:
NAME/KEY: misc feature
LOCATION: 165..6722
OTHER INFORMATION: /function= "Nucleotides 165-6722
encode the mature peptide"
OTHER INFORMATION: /note= "Identification Method: S"
US-08-727-034-6

Query Match 68.2%; Score 15; DB 1; Length 6843;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGGCGAAAACCTCCG 16
|||||
Db 5908 CGGCGAAAACCTCCG 5922

RESULT 25
US-08-858-207A-177
Sequence 177, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Scodola, Robert
TITLE OF INVENTION: NO. 6348328el Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 177:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-858-207A-177

Query Match 67.3%; Score 14.8; DB 4; Length 774;
Best Local Similarity 88.9%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCAAAACCTCCGCCTAAC 22
| | | | | | | | | |
Db 470 GCAAAACCTCCGCCTGAC 487

RESULT 26

US-09-328-352-715/c
; Sequence 715, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 715
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-715

Query Match 67.3%; Score 14.8; DB 4; Length 1038;
Best Local Similarity 88.9%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GCGAAACCTCCGCCTAA 21
| | | | | | | | | |
Db 366 GCGAAACCTCCGCCTAA 349

RESULT 27

US-09-489-039A-1114/c
; Sequence 1114, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1114
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1114

Query Match 67.3%; Score 14.8; DB 4; Length 1206;
Best Local Similarity 88.9%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCCGCCTA 20
| | | | | | | | | |
Db 577 GGGCAAAACCTCCGCCTA 560

RESULT 28

US-09-543-681A-445/c
; Sequence 445, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 445
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-445

Query Match 67.3%; Score 14.8; DB 4; Length 1482;
Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GCGAAACCTCCGCCTAA 21
| | | | | | | | | |
Db 105 GCGTAACCTCCGCCTAA 88

RESULT 29

US-09-620-312D-723/c
; Sequence 723, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, fuding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 723
; LENGTH: 1660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (127)..(1401)
US-09-620-312D-723

Query Match 67.3%; Score 14.8; DB 4; Length 1660;
Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GCGAAACCTCCGCCTAA 21
| | | | | | | | | |
Db 363 GCGTAACCTCCGCCTAA 346

RESULT 30

US-09-489-039A-1224
; Sequence 1224, Application US/09489039A
; Patent No. 6610836

; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1224
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1224

Query Match 67.3%; Score 14.8; DB 4; Length 1770;
Best Local Similarity 88.9%; Pred. No. 2.6e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCGCCTA 20
| | | | | | | | | | | | | | | | | | | | | |
Db 1194 GGGCAAAACCAACCCGCGA 1211

Search completed: June 20, 2004, 11:46:26
Job time : 40.453 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 124.434 Seconds
(without alignments)
809.940 Million cell updates/sec

Title: US-10-624-714-14

Perfect score: 22

Sequence: 1 ccgggcaaaacctccgcctaac 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA:
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:
9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:
10: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	95.5	723	9	US-09-712-363-114
2	17.8	80.9	2629	16	US-10-348-052-24
3	17.2	78.2	1064	13	US-10-424-599-66751
4	17.2	78.2	5925	15	US-10-311-455-1549
5	16.8	76.4	416	13	US-10-424-599-31632
6	16.8	76.4	536	11	US-09-864-408A-265
7	16.8	76.4	1728	16	US-10-260-238-1137
8	16.4	74.5	927	9	US-09-815-242-7429
9	16.4	74.5	933	13	US-10-335-977-3347
10	16.4	74.5	1035	13	US-10-335-977-3348
11	16.4	74.5	1320	13	US-10-425-114-6455
12	16.4	74.5	1320	13	US-10-282-122A-36715
13	16.4	74.5	1385	13	US-10-425-114-29496
14	16.4	74.5	1492	13	US-10-424-599-102731
15	15.4	70.0	15.4	70.0	Sequence 114, App
16	15.4	70.0	15.4	70.0	Sequence 24, Appl
17	15.4	70.0	15.4	70.0	Sequence 66751, A
18	15.4	70.0	15.4	70.0	Sequence 1549, Ap
19	15.4	70.0	15.4	70.0	Sequence 31632, A
20	15.4	70.0	15.4	70.0	Sequence 265, App
21	15.4	70.0	15.4	70.0	Sequence 1137, Ap
22	15.4	70.0	15.4	70.0	Sequence 7429, Ap
23	15.4	70.0	15.4	70.0	Sequence 3347, Ap
24	15.4	70.0	15.4	70.0	Sequence 3348, Ap
25	15.4	70.0	15.4	70.0	Sequence 6455, Ap
26	15.4	70.0	15.4	70.0	Sequence 36715, A
27	15.4	70.0	15.4	70.0	Sequence 29496, A
28	15.4	70.0	15.4	70.0	Sequence 102731,

15	16.4	74.5	3832	15	US-10-102-524-1718	Sequence 1718, Ap
16	16.4	74.5	3832	15	US-10-376-564-86	Sequence 86, Appl
17	16.4	74.5	3832	16	US-10-159-563-291	Sequence 291, App
18	16.2	73.6	734	13	US-10-027-632-144898	Sequence 144898,
19	16.2	73.6	734	13	US-10-027-632-144898	Sequence 144898,
20	16.2	73.6	753	12	US-09-727-892-10	Sequence 10, Appl
21	16.2	73.6	1911	12	US-09-973-674A-33	Sequence 33, Appl
22	16.2	73.6	6022	13	US-10-221-714A-119	Sequence 119, App
23	16.2	73.6	6022	15	US-10-239-676-79	Sequence 79, Appl
24	16.2	73.6	6022	15	US-10-311-455-1045	Sequence 1045, Ap
25	16.2	73.6	6022	15	US-10-240-455-85	Sequence 85, Appl
26	16.2	73.6	9888	15	US-10-311-455-1214	Sequence 1214, Ap
27	16.2	73.6	16668	12	US-09-727-892-1	Sequence 1, Appli
28	15.8	71.8	158	13	US-10-065-783A-15488	Sequence 15488, A
29	15.8	71.8	158	16	US-10-242-535A-15488	Sequence 15488, A
30	15.8	71.8	346	15	US-10-106-698-1566	Sequence 1566, Ap
31	15.8	71.8	611	13	US-10-424-599-31840	Sequence 31840, A
32	15.8	71.8	990	16	US-10-260-238-4066	Sequence 4066, Ap
33	15.8	71.8	1019	13	US-10-425-114-27844	Sequence 27844, A
34	15.8	71.8	1101	9	US-09-972-186A-3	Sequence 3, Appli
35	15.8	71.8	1375	13	US-10-183-687-466	Sequence 466, App
36	15.8	71.8	1375	15	US-10-180-375-213	Sequence 213, App
37	15.8	71.8	1403	16	US-10-369-493-38599	Sequence 38599, A
38	15.8	71.8	1415	16	US-10-369-493-35378	Sequence 35378, A
39	15.8	71.8	1415	16	US-10-369-493-38195	Sequence 38195, A
40	15.8	71.8	2307	13	US-10-282-122A-12392	Sequence 12392, A
41	15.8	71.8	22507	15	US-10-085-959-53	Sequence 53, Appl
42	15.6	70.9	117	9	US-09-867-701-8023	Sequence 8023, Ap
43	15.6	70.9	376	13	US-10-424-599-52390	Sequence 52390, A
44	15.6	70.9	403	9	US-09-960-352-3319	Sequence 3319, Ap
45	15.6	70.9	473	16	US-10-260-238-3779	Sequence 3779, Ap
46	15.6	70.9	558	13	US-10-282-122A-40260	Sequence 40260, A
47	15.6	70.9	957	13	US-10-282-122A-26820	Sequence 26820, A
48	15.6	70.9	1041	10	US-09-953-348-38	Sequence 38, Appl
49	15.6	70.9	1041	15	US-10-267-255-38	Sequence 38, Appl
50	15.6	70.9	1150	16	US-10-260-238-3780	Sequence 3780, Ap
51	15.6	70.9	1182	13	US-10-282-122A-23981	Sequence 23981, A
52	15.6	70.9	1417	17	US-10-426-002-21	Sequence 21, Appl
53	15.6	70.9	1915	13	US-10-027-632-98082	Sequence 98082, A
54	15.6	70.9	1915	13	US-10-027-632-98325	Sequence 98325, A
55	15.6	70.9	1915	16	US-10-027-632-98082	Sequence 98082, A
56	15.6	70.9	2400	16	US-10-369-493-37751	Sequence 37751, A
57	15.6	70.9	2497	13	US-10-424-599-14892	Sequence 14892, A
58	15.6	70.9	5663	15	US-10-311-455-670	Sequence 670, App
59	15.6	70.9	5663	17	US-10-240-589C-32	Sequence 32, Appl
60	15.6	70.9	7029	9	US-09-826-191-1	Sequence 1, Appli
61	15.6	70.9	7029	15	US-10-263-666-1	Sequence 1, Appli
62	15.6	70.9	7029	15	US-10-261-481-1	Sequence 1, Appli
63	15.6	70.9	7029	15	US-10-261-942-1	Sequence 1, Appli
64	15.6	70.9	8115	9	US-09-778-516A-1	Sequence 1, Appli
65	15.6	70.9	8115	9	US-09-778-516A-2	Sequence 2, Appli
66	15.6	70.9	10183	13	US-10-221-714A-478	Sequence 478, App
67	15.6	70.9	19513	13	US-10-221-613-206	Sequence 206, App
68	15.6	70.9	19513	15	US-10-284-708-40	Sequence 40, Appl
69	15.6	70.9	31940	13	US-10-672-787-13	Sequence 13, Appl
70	15.6	70.9	43411	17	US-10-450-826-76	Sequence 76, Appl
71	15.6	70.9	53500	10	US-09-953-348-76	Sequence 76, Appl
72	15.6	70.9	53500	15	US-10-267-255-76	Sequence 106, App
73	15.6	70.9	96593	12	US-10-052-482-106	Sequence 43, Appl
74	15.4	70.0	514	9	US-09-833-790-43	Sequence 274118,
75	15.4	70.0	514	13	US-10-027-632-274118	Sequence 274118,
76	15.4	70.0	548	13	US-10-027-632-274118	Sequence 45, Appl
77	15.4	70.0	548	16	US-09-811-284-45	Sequence 105614,
78	15.4	70.0	616	9	US-10-027-632-105614	Sequence 105614,
79	15.4	70.0	674	13	US-10-027-632-105614	Sequence 101014,
80	15.4	70.0	674	16	US-10-027-632-101014	Sequence 101014,
81	15.4	70.0	1066	13	US-10-027-632-101014	Sequence 26769, A
82	15.4	70.0	1066	16	US-10-369-493-26769	Sequence 26769, A
83	15.4	70.0	1149	16	US-09-974-300-2965	Sequence 2965, Ap
84	15.4	70.0	1233	9	US-09-815-242-5924	Sequence 5924, Ap
85	15.4	70.0	1401	9	US-10-282-122A-20221	Sequence 20221, A
86	15.4	70.0	1401	13	US-10-425-114-12027	Sequence 12027, A
87	15.4	70.0	2059	13		

Sequence 179, App
Sequence 32500, A
Sequence 47225, A
Sequence 47225, A
Sequence 25778, A
Sequence 18289, A
Sequence 18289, A
Sequence 18289, A
Sequence 4489, App
Sequence 61, Appl
Sequence 4489, App
Sequence 1320, A
Sequence 266933, A
Sequence 266933, A
Sequence 23806, A
Sequence 23806, A
Sequence 1384, App
Sequence 14952, A
Sequence 1307, App
Sequence 12781, A
Sequence 243647, A
Sequence 1663, App
Sequence 21935, A
Sequence 25, Appl
Sequence 28574, A
Sequence 28591, A
Sequence 28591, A
Sequence 40993, A
Sequence 28880, A
Sequence 28744, A
Sequence 126744, A
Sequence 126744, A
Sequence 60, Appl
Sequence 122263, A
Sequence 122263, A
Sequence 122263, A
Sequence 122264, A
Sequence 77, Appl
Sequence 35, Appl
Sequence 2063, App
Sequence 26589, A
Sequence 1, Appl
Sequence 28142, A
Sequence 107, App
Sequence 43, Appl
Sequence 205, App
Sequence 32492, A
Sequence 700, App
Sequence 401, App
Sequence 183, App
Sequence 183, App
Sequence 342, App
Sequence 341, App
Sequence 278, App
Sequence 57, Appl
Sequence 688, App
Sequence 30755, A
Sequence 107, App
Sequence 13424, A
Sequence 1807, App
Sequence 41, Appl
Sequence 77, Appl
Sequence 2372, App

US-09-918-995-179
US-09-918-995-32500
US-10-085-783A-47225
US-10-085-783A-47225
US-09-918-995-25778
US-10-085-783A-18289
US-10-085-783A-18289
US-09-796-692-4489
US-10-040-862-4489
US-10-175-499-61
US-10-057-475B-4489
US-10-154-84B-4489
US-10-282-122A-19320
US-10-027-632-266933
US-10-027-632-266933
US-10-027-632-29806
US-10-027-632-29806
US-10-066-543-1384
US-10-085-783A-14952
US-10-242-535A-14952
US-10-425-114-1307
US-10-198-846-12781
US-10-027-632-243647
US-10-027-632-243647
US-10-264-049-1663
US-10-425-114-21935
US-10-231-417-45
US-10-027-632-26574
US-10-027-632-26574
US-10-027-632-28591
US-10-027-632-28591
US-10-369-493-40993
US-10-282-122A-28880
US-10-282-122A-29744
US-10-027-632-126744
US-10-027-632-126744
US-10-175-499-60
US-10-027-632-122263
US-10-027-632-122264
US-10-027-632-122263
US-10-027-632-122264
US-10-084-846A-77
US-10-330-051A-35
US-10-156-761-2063
US-10-282-122A-26589
US-09-712-363-1
US-10-282-122A-28142
US-10-403-571-107
US-10-225-066A-43
US-10-374-780A-205
US-10-425-114-32492
US-09-738-626-700
US-10-106-698-401
US-09-925-299-183
US-09-925-299-183
US-09-919-039-342
US-09-919-039-341
US-10-116-275-278
US-10-288-930-57
US-09-738-626-688
US-10-282-122A-30755
US-10-363-616-107
US-10-198-846-13424
US-10-062-674-1807
US-10-197-666A-77
US-10-252-157-41
US-10-108-260A-2372

ALIGNMENTS

RESULT 1

US-09-712-363-114
; Sequence 114, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-114

Query Match 95.5%; Score 21; DB 9; Length 723;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCCTAA 21
DB 703 CCGGCAAAACCTCGCCTAA 723

RESULT 2
US-10-348-052-24/c
; Sequence 24, Application US/10348052
; Publication No. US20030219782A1
; GENERAL INFORMATION:
; APPLICANT: Saba, Julie D.
; APPLICANT: Fyrist, Henrik
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE MODULATION
; TITLE OF INVENTION: OF SPHINGOLIPID METABOLISM AND/OR SIGNALING
; FILE REFERENCE: 200116.405
; CURRENT APPLICATION NUMBER: US/10/348,052
; CURRENT FILING DATE: 2003-01-17
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 2629
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-348-052-24

Query Match 80.9%; Score 17.8; DB 16; Length 2629;
Best Local Similarity 90.5%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCGGCAAAACCTCGCCTAA 21

D_b 2126 CCGGGCATAACTCCGCCTGA 2106

RESULT 3

```

US_10-424-599-66751
; Sequence 66751, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Mole
; TITLE OF INVENTION: Plants and Uses Ther
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 66751
; LENGTH: 1064
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847
US_10-424-599-66751

```

Query Match	78.2%;	Score 17.2;	DB 13;	Length 1064;
Best Local Similarity	86.4%;	Pred. No. 1.2e+02;		
Matches 19; Conservative	0;	Mismatches 3;	Indels 0;	

Qy 1 CCGGCGAAACCTCGCCTAAC 22
Dδ 247 CCGGCGAAACCTCGCCTAAC 268

RESULT 4

```

US-10-311-455-1549/c
; Sequence 1549, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10035259.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1549
; LENGTH: 5925
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated gene
US-10-311-455-1549

```

```
Query Match          78.2%; Score 17.2; DB 15; Length 5925;
Best Local Similarity 86.4%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0;
```

QY
db

1 CCGGGCAAAACCTTCGCGCTAAC 22
||| ||| ||| ||| ||| ||| |||
4911 CCGGGCAAAACCGCCCTAAC 489

RESULT 5

```

US-10-424-599-31632/c
; Sequence 31632, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhao Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid M
; TITLE OF INVENTION: Plants and Uses T
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 295684
; SEQ ID NO 31632
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3
US-10-424-599-31632

```

Query Match	76.4%	Score 16.8;	DB 13;	Length 416;
Best Local Similarity	90.0%;	Pred. No. 1.9e+02;		
Matches 18: Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY 3 GGGCAAAACCTCCGCCCTAAC 22
|||
Db 318 GGGCAAAACCAACCGCCCAAC 299

RESIT.T. 6

```

US-09-864-408A-265/c
; Sequence 265, Application US/09864408A
; Publication NO. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No.US20040009474A1el Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 265
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-864-408A-265

```

Query Match 76.4%; Score 16.8; DB 11; Length 536;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCGCCTAAC 22
Db 196 GGGCAAAAGCTGCGCCTAAC 177

RESULT 7

```

US-102-260-238-1137
; Sequence 1137, Application US/10260238
; Publication No. US2004001602SA1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyuki

```

APPLICANT: Kreps, Joel
APPLICANT: Provart, Nicholas
APPLICANT: Riche, Darrell
APPLICANT: Zhu, Tong
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
FILE REFERENCE: 60111-NP
CURRENT APPLICATION NUMBER: US/10/260,238
CURRENT FILING DATE: 2002-09-26
PRIOR APPLICATION NUMBER: US 60/325,448
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 1137
LENGTH: 1728
TYPE: DNA
ORGANISM: Oryza sativa
US-10-260-238-1137

Query Match 76.4%; Score 16.8; DB 16; Length 1728;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGCAGAAACCTCGGCTAAC 22
DB 337 GGCAGAAACCTCGGCTAAC 356

RESULT 8
US-09-815-242-7429/c
Sequence 7429, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Chlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7429
LENGTH: 927
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (1)....(927)
US-09-815-242-7429

Query Match 74.5%; Score 16.4; DB 9; Length 927;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCAGAAACCTCGGCTAAC 21
DB 249 GGCAGAAACCTCGGCTAAC 232

RESULT 9
US-10-335-977-3347/c
Sequence 3347, Application US/10335977
Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 3347:
SEQUENCE CHARACTERISTICS:
LENGTH: 927 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...927
SEQUENCE DESCRIPTION: SEQ ID NO: 3347:
US-10-335-977-3347

Query Match 74.5%; Score 16.4; DB 13; Length 927;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCAGAAACCTCGGCTAAC 21
DB 249 GGCAGAAACCTCGGCTAAC 232

RESULT 10
US-10-335-977-3348/c
Sequence 3348, Application US/10335977

Publication No. US20040052799A1
GENERAL INFORMATION:
APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 3348:
SEQUENCE CHARACTERISTICS:
LENGTH: 933 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...933
SEQUENCE DESCRIPTION: SEQ ID NO: 3348:
US-10-335-977-3348
Query Match 74.5%; Score 16.4; DB 13; Length 933;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 GGCAAAACCTCCGCCTAA 21
DB 258 GGCAAAACCTCCGCCTAA 241
RESULT 11
US-10-425-114-6455
Sequence 6455, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114

US-10-425-114-6455
Sequence 6455, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114

Query Match 74.5%; Score 16.4; DB 13; Length 1035;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 GGCAAAACCTCCGCCTAA 21
DB 99 GGCAAAACCTCCGCCTAA 116

RESULT 12
US-10-282-122A-36715/C
Sequence 36715, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 36715
LENGTH: 1320
TYPE: DNA
ORGANISM: Salmonella paratyphi A
US-10-282-122A-36715

Query Match 74.5%; Score 16.4; DB 13; Length 1320;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGCAAAACCTCCGCC 18
Db 1237 CCAGGCAAAACCTCCGCC 1220

RESULT 13
US-10-425-114-29496
; Sequence 29496, Application US/10425114
; Publication No. US20040034988A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 29496
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-GMNONOIR024E05_FLI
US-10-425-114-29496

Query Match 74.5%; Score 16.4; DB 13; Length 1385;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCAAAACCTCCGCTAA 21
Db 335 GGCAAAACCTCCCTCTAA 352

RESULT 14
US-10-424-599-102731
; Sequence 102731, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 102731
; LENGTH: 1492
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1492)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_63784C.1
US-10-424-599-102731

Query Match 74.5%; Score 16.4; DB 13; Length 1492;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGCAAAACCTCCGCTAA 21
Db 353 GGCAAAACCTCCCTCTAA 370

RESULT 15
US-10-102-524-1718
; Sequence 1718, Application US/10102524
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF KIDNEY CANCER
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1718
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-102-524-1718

Query Match 74.5%; Score 16.4; DB 15; Length 3832;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCCT 19
Db 1164 CGGGCAAAACCTCCACCT 1181

RESULT 16
US-10-376-564-86
; Sequence 86, Application US/10376564
; Publication No. US20030180302A1
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckhard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014003
; CURRENT APPLICATION NUMBER: US/10/376,564
; CURRENT FILING DATE: 2003-02-29
; PRIOR APPLICATION NUMBER: US 09/886,319
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 100 30 149,5-41
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 86
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-376-564-86

Query Match 74.5%; Score 16.4; DB 15; Length 3832;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CGGGCAAAACCTCCGCCT 19
Db 1164 CGGGCAAAACCTCCACCT 1181

```
RESULT 17
US-10-159-563-291
; Sequence 291, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Metzger, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
; FILE REFERENCE: 11613.56US11
; CURRENT APPLICATION NUMBER: US/10/159,563
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: US 10/133,937
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 291
; LENGTH: 3832
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-291

Query Match      74.5%; Score 16.4; DB 16; Length 3832;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CGGGCAAAACCTCCGCCT 19
DB      1164 CGGGCAAAACCTCCACCT 1181

RESULT 18
US-10-027-632-144898/c
; Sequence 144898, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144898
; LENGTH: 734
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144898

Query Match      73.6%; Score 16.2; DB 13; Length 734;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CGGGCAAAACCTCCGCCTAA 21
DB      1 CGGGCAAAACCTCCGCCTAA 21

RESULT 19
US-10-027-632-144898/c
; Sequence 144898, Application US/10027632
; Publication No. US20030204075A3
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144898
; LENGTH: 734
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144898

Query Match      73.6%; Score 16.2; DB 16; Length 734;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CGGGCAAAACCTCCGCCTAA 21
DB      515 CTGGCAAAACCTAAGCCTAA 495

RESULT 20
US-09-727-892-10/c
; Sequence 10, Application US/0927892
; Publication No. US20040091856A1
; GENERAL INFORMATION:
; APPLICANT: Phagotech, Inc.
; APPLICANT: BELLETIER, Jerry
; APPLICANT: GROS, Philippe
; APPLICANT: DUBOW, Michael
; TITLE OF INVENTION: DATA SEQUENCES FROM STAPHYLOCOCCUS AUREUS BACTERIOPHAGE 44 AHJD
; TITLE OF INVENTION: THAT ENCODE ANTI-MICROBIAL POLYPEPTIDES
; FILE REFERENCE: 073406-0302
; CURRENT APPLICATION NUMBER: US/09/727,892
; CURRENT FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 10
; LENGTH: 753
; TYPE: DNA
; ORGANISM: Staphylococcus aureus Bacteriophage 44 AHJD
US-09-727-892-10

Query Match      73.6%; Score 16.2; DB 12; Length 753;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CGGGCAAAACCTCCGCCTAAC 22
DB      2 CGGGCAAAACCTCCGCCTAAC 22
```

Db 344 CGGTCAAAACCGCGCTAAC 324

RESULT 21

US-09-973-674A-33

; Sequence 33, Application US/09973674A

; Publication No. US20040086937A1

; GENERAL INFORMATION:

; APPLICANT: Loferer, H.

; APPLICANT: Jacobi, A.

; TITLE OF INVENTION: NOVEL METHOD FOR IDENTIFYING ANTIBACTERIAL COMPOUNDS

; FILE REFERENCE: GPCG-P01-019

; CURRENT APPLICATION NUMBER: US/09/973,674A

; PRIOR FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: PCT/EP00/03135

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: EP 00102111.2

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: EP 99107031.9

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 33

; LENGTH: 1911

; TYPE: DNA

; ORGANISM: Escherichia coli

US-09-973-674A-33

Query Match 73.6%; Score 16.2; DB 12; Length 1911;

Best Local Similarity 85.7%; Pred. No. 3.6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCGGCTAAC 22

Db 146 CGGCAAAACCTACGCTTACC 166

RESULT 22

US-10-221-714A-119/c

; Sequence 119, Application US/10221714A

; Publication No. US20040048254A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with

; FILE REFERENCE: 5013.1005

; CURRENT APPLICATION NUMBER: US/10/221,714A

; PRIOR FILING DATE: 2003-01-21

; PRIOR APPLICATION NUMBER: PCT/EP01/02955

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: DE 10013847.0

; PRIOR FILING DATE: 2000-03-15

; PRIOR APPLICATION NUMBER: DE 10019058.8

; PRIOR FILING DATE: 2000-04-06

; PRIOR APPLICATION NUMBER: DE 10019173.8

; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 540

; SEQ ID NO 119

; LENGTH: 6022

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-221-714A-119

Query Match 73.6%; Score 16.2; DB 13; Length 6022;

Best Local Similarity 85.7%; Pred. No. 3.6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCGGCTAAC 22

Db 5377 CGGCAAAACGTCCGCCGAAC 5357

RESULT 23

US-10-239-676-79/c

; Sequence 79, Application US/10239676

; Publication No. US20030082609A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

; FILE REFERENCE: 5013.1003

; CURRENT APPLICATION NUMBER: US/10/239,676

; PRIOR FILING DATE: 2002-09-24

; PRIOR APPLICATION NUMBER: PCT/EP01/03968

; DE 10019058.8

; DE 10019173.8

; DE 10032529.7

; DE 10043826.1

; PRIOR FILING DATE: 2001-04-06

; 2000-04-06

; 2000-04-07

; 2000-06-30

; 2000-09-01

; NUMBER OF SEQ ID NOS: 228

; SEQ ID NO 79

; LENGTH: 6022

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-239-676-79

Query Match 73.6%; Score 16.2; DB 15; Length 6022;

Best Local Similarity 85.7%; Pred. No. 3.6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCGGCTAAC 22

Db 5377 CGGCAAAACGTCCGCCGAAC 5357

RESULT 24

US-10-311-455-1045/c

; Sequence 1045, Application US/10311455

; Publication No. US20030143606A1

; GENERAL INFORMATION:

; APPLICANT: OLEK, Alexander

; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt

; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ

; FILE REFERENCE: 5013.1014

; CURRENT APPLICATION NUMBER: US/10/311,455

; PRIOR FILING DATE: 2002-12-16

; PRIOR APPLICATION NUMBER: PCT/EP01/07537

; PRIOR FILING DATE: 2001-07-02

; PRIOR APPLICATION NUMBER: DE 10032529.7

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: DE 10043826.1

; PRIOR FILING DATE: 2000-09-01

; NUMBER OF SEQ ID NOS: 2424

; SEQ ID NO 1045

; LENGTH: 6022

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-311-455-1045

Query Match 73.6%; Score 16.2; DB 15; Length 6022;
 Best Local Similarity 85.7%; Pred. No. 3.6e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCTAAC 22
 |||||
 DB 5377 CGGCAAAACGTCGCGCAAC 5357

RESULT 25

US-10-240-453-85/c

; Sequence 85, Application US/10240453
 ; Publication No. US20030148326A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
 ; TITLE OF INVENTION: transcription
 ; TITLE OF INVENTION: By Means of Assessing the Methylation Status of Genes Associated
 ; TITLE OF INVENTION: With DNA Transcription
 ; FILE REFERENCE: 5013.1009
 ; CURRENT APPLICATION NUMBER: US/10/240,453
 ; CURRENT FILING DATE: 2002-10-02
 ; PRIOR APPLICATION NUMBER: PCT/EP01/03973
 ; PRIOR FILING DATE: 2001-04-06
 ; PRIOR APPLICATION NUMBER: DE 10019058.8
 ; PRIOR FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: DE 10019173.8
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01
 ; NUMBER OF SEQ ID NOS: 350
 ; SEQ ID NO 85
 ; LENGTH: 6022
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-240-453-85

Query Match 73.6%; Score 16.2; DB 15; Length 6022;
 Best Local Similarity 85.7%; Pred. No. 3.6e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCTAAC 22
 |||||
 DB 5377 CGGCAAAACGTCGCGCAAC 5357

RESULT 26

US-10-311-455-1214/c

; Sequence 1214, Application US/10311455
 ; Publication No. US20030143606A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OLEK, Alexander
 ; APPLICANT: PIEPENBROCK, Christian
 ; APPLICANT: BERLIN, Kurt
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
 ; TITLE OF INVENTION: cytosine methylation
 ; FILE REFERENCE: 5013.1014
 ; CURRENT APPLICATION NUMBER: US/10/311,455
 ; CURRENT FILING DATE: 2002-12-16
 ; PRIOR APPLICATION NUMBER: PCT/EP01/07537
 ; PRIOR FILING DATE: 2001-07-02
 ; PRIOR APPLICATION NUMBER: DE 10032529.7
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: DE 10043826.1
 ; PRIOR FILING DATE: 2000-09-01

US-10-311-455-1214

Query Match 73.6%; Score 16.2; DB 15; Length 9888;
 Best Local Similarity 85.7%; Pred. No. 3.6e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCTAAC 22
 |||||
 DB 5084 CGGCAAAACCTCCGCTAAC 5064

RESULT 27

US-09-727-892-1/c

; Sequence 1, Application US/09727892
 ; Publication No. US20040091856A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Phagotech, Inc.
 ; APPLICANT: PELLETIER, Jerry
 ; APPLICANT: GROS, Philippe
 ; APPLICANT: DUBOW, Michael
 ; TITLE OF INVENTION: DNA SEQUENCES FROM STAPHYLOCOCCUS AUREUS BACTERIOPHAGE 44 AHJD
 ; TITLE OF INVENTION: THAT ENCODE ANTI-MICROBIAL POLYPEPTIDES
 ; FILE REFERENCE: 073406-0302
 ; CURRENT APPLICATION NUMBER: US/09/727,892
 ; CURRENT FILING DATE: 2000-12-01
 ; NUMBER OF SEQ ID NOS: 159
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 16668
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus aureus Bacteriophage 44 AHJD

US-09-727-892-1

Query Match 73.6%; Score 16.2; DB 12; Length 16668;
 Best Local Similarity 85.7%; Pred. No. 3.5e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGGCAAAACCTCCGCTAAC 22
 |||||
 DB 6087 CGGCAAAACCTCCGCTAAC 6067

RESULT 28

US-10-085-783A-15488

; Sequence 15488, Application US/10085783A
 ; Publication No. US20040037841A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; APPLICANT: Liew, C.C.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2002
 ; CURRENT APPLICATION NUMBER: US/10/085,783A
 ; CURRENT FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: US 60/275,017
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 60/271,955
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 58994
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 15488
 ; LENGTH: 158
 ; TYPE: DNA
 ; ORGANISM: Human
 ; ORGANISM: US-10-085-783A-15488

Query Match 71.8%; Score 15.8; DB 13; Length 158;
Best Local Similarity 89.5%; Pred. No. 5.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCGCCTAA 21
|||||
Db 127 GGGCAAAACCTCGACCTAA 145

RESULT 29

US-10-242-535A-15488
; Sequence 15488, Application US/10242535A
; Publication No. US20040013663A1

; GENERAL INFORMATION:

; APPLICANT: ChondroGene Inc.

; APPLICANT: Liew, C.C.

; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis

; FILE REFERENCE: 4231/2005

; CURRENT APPLICATION NUMBER: US/10/242,535A

; CURRENT FILING DATE: 2002-09-12

; PRIOR APPLICATION NUMBER: US 10/085,783

; PRIOR FILING DATE: 2002-02-28

; PRIOR APPLICATION NUMBER: US 60/305,340

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/275,017

; PRIOR FILING DATE: 2001-03-12

; PRIOR APPLICATION NUMBER: US 60/271,955

; PRIOR FILING DATE: 2001-02-28

; NUMBER OF SEQ ID NOS: 58994

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 15488

; LENGTH: 158

; TYPE: DNA

; ORGANISM: Human

US-10-242-535A-15488

Query Match 71.8%; Score 15.8; DB 16; Length 158;
Best Local Similarity 89.5%; Pred. No. 5.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGCAAAACCTCGCCTAA 21
|||||
Db 127 GGGCAAAACCTCGACCTAA 145

RESULT 30

US-10-106-698-1566/c

; Sequence 1566, Application US/10106698

; Publication No. US20030105690A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

; FILE REFERENCE: PA005P1

; CURRENT APPLICATION NUMBER: US/10/106,698

; CURRENT FILING DATE: 2002-03-27

; PRIOR APPLICATION NUMBER: PCT/US00/26524

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING DATE: 1999-09-29

; PRIOR APPLICATION NUMBER: US 60/163,280

; PRIOR FILING DATE: 1999-11-03

; NUMBER OF SEQ ID NOS: 8564

; SOFTWARE: PatentIn Ver. 3.0

; SEQ ID NO 1566

; LENGTH: 346

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (312)..(312)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc_feature

; LOCATION: (314)..(314)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1566

Query Match 71.8%; Score 15.8; DB 15; Length 346;
Best Local Similarity 81.0%; Pred. No. 5.7e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGGCAAAACCTCGCCTAA 21
|||||
Db 319 CCGGGNANAACTGGCCAAA 299

Search completed: June 20, 2004, 17:45:58
Job time : 136.434 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 705.989 Seconds
(without alignments)
1596.226 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26

Sequence: 1 acaacggagggccgaatctacgaa 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.pl.*

8: gb.pr.*

9: gb.ro.*

10: gb.sts.*

11: gb.un.*

12: gb.un.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.in.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pin.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	26	100.0	401	6	AR149059	Sequence
C 2	26	100.0	401	6	AR340431	Sequence
C 3	26	100.0	18121	1	AE007182	Mycobacte
C 4	26	100.0	37164	1	MSGY126	Mycobacte
C 5	26	100.0	244800	1	EX842584	Mycobacte
C 6	26	100.0	278492	1	EX248347	Mycobacte
C 7	19.6	75.4	12461	1	AE007304	Sinorhizo
C 8	19.4	74.6	38793	1	MSGB27CS	Mycobacteri
C 9	19.4	74.6	348450	1	MLEPRTN4	Mycobacte
C 10	18.6	71.5	505	6	AR226916	Sequence
C 11	18.6	71.5	1236	8	BT010447	Arabidops
C 12	18.6	71.5	1766	3	BT011075	Drosophil
C 13	18.6	71.5	1959	9	AB046101	Macaca fa
C 14	18.6	71.5	2057	6	BD180518	Highly th
C 15	18.6	71.5	2358	3	AF301149	Drosophil
C 16	18.6	71.5	40970	3	LMFLUNK03	Leishmani
C 17	18.6	71.5	42950	2	AC017471	Drosophil
C 18	18.6	71.5	103904	8	ATAC021640	Arabidops
C 19	18.6	71.5	110000	2	LMFLCHR31_13	Continuation (14 o
C 20	18.6	71.5	110000	2	LMFLCHR32_17	Continuation (18 o
C 21	18.6	71.5	158402	3	AC008343	Drosophil
C 22	18.6	71.5	164840	2	AC011127	Homo sapi
C 23	18.6	71.5	18553	9	AC022046	Homo sapi
C 24	18.6	71.5	189412	2	AC023532	Homo sapi
C 25	18.6	71.5	197837	2	AP001397	Homo sapi
C 26	18.6	71.5	205152	9	AP002985	Homo sapi
C 27	18.6	71.5	214701	2	AP001447	Homo sapi
C 28	18.6	71.5	255652	3	AE003811	Drosophil
C 29	18.6	71.5	301618	1	AE017177	Porphyrom
C 30	18.6	71.5	344615	1	EX569695	Synechoco
C 31	18.2	70.0	163025	14	AY372243	Pstittaci
C 32	18.2	70.0	176539	9	AL354928	Human DNA
C 33	18	69.2	18857	1	AE006947	Mycobacte
C 34	18	69.2	22913	2	AC130497	Rattus no
C 35	18	69.2	262082	2	AC120441	Rattus no
C 36	18	69.2	315876	2	AC112779	Rattus no
C 37	18	69.2	324050	1	EX248335	Mycobacte
C 38	18	69.2	325841	2	AC094665	Rattus no
C 39	18	69.2	337425	2	AC105666	Rattus no
C 40	18	69.2	337436	1	AC098285	Rattus no
C 41	18	69.2	342416	1	EX842573	Mycobacte
C 42	18	69.2	343250	1	AP003594	Nostoc sp
C 43	17.8	68.5	298700	1	AP005951	Bradyrhiz
C 44	17.6	67.7	9973	1	AE000070	Rhizobium
C 45	17.6	67.7	11159	1	AE012066	Xanthomon
C 46	17.6	67.7	38141	2	AC018202	Drosophil
C 47	17.6	67.7	82849	3	AC004277	Drosophil
C 48	17.6	67.7	100589	8	OSJN00191	Oryza sat
C 49	17.6	67.7	110000	2	LMFLCHR31_13	Continuation (14 o
C 50	17.6	67.7	110000	2	LMFLCHR31_14	Continuation (15 o
C 51	17.6	67.7	137332	8	AP002820	Oryza sat
C 52	17.6	67.7	143710	8	AP002899	Oryza sat
C 53	17.6	67.7	146840	8	AP003343	Oryza sat
C 54	17.6	67.7	151203	8	CNS09S4W	Oryza sat
C 55	17.6	67.7	153064	8	CNS08CAL	Oryza sat
C 56	17.6	67.7	165512	3	AC099010	Drosophil
C 57	17.6	67.7	175667	9	AC009336	Homo sapi
C 58	17.6	67.7	245560	1	AP005330	Vibrio vu
C 59	17.6	67.7	270766	3	AE003615	Drosophil
C 60	17.6	67.7	298550	1	AP005961	Bradyrhiz
C 61	17.6	67.7	300880	1	AE016917	Chromobac
C 62	17.6	67.7	301235	1	AE016799	Vibrio vu
C 63	17.6	67.7	301838	1	AE017209	Geobacter
C 64	17.6	67.7	314250	1	EX294149	Pirellula
C 65	17.6	67.7	320040	6	A79350	Sequence 1

C 66	17.6	67.7	320040	6	A93002	A93002 Sequence 1	C 139	17	65.4	146500	2	AC119500	Leishmani
C 67	17.6	67.7	348971	1	BX572594	Rhodopsin	140	17	65.4	147844	2	AC068188	Homo sapi
C 68	17.4	66.9	1970	8	AY163775	Arabidops	141	17	65.4	154541	8	AP003278	Oryza sat
C 69	17.4	66.9	1982	8	AY163773	Arabidops	142	17	65.4	159843	5	EX000485	Zebrafish
C 70	17.4	66.9	2004	8	AY163774	Arabidops	143	17	65.4	163671	8	AP003330	Oryza sat
C 71	17.2	66.2	13230	1	AE014659	Bifidobac	144	17	65.4	166757	8	AC098565	Oryza sat
C 72	17.2	66.2	13632	1	AE000851	Methanoba	145	17	65.4	166757	8	AC098565	Oryza sat
C 73	17.2	66.2	161787	1	AC105679	Rattus no	C 146	17	65.4	170523	9	AP002387	Homo sapi
C 74	17.2	66.2	223820	2	AC094762	Rattus no	147	17	65.4	171415	2	AP002771	Homo sapi
C 75	17.2	66.2	225245	2	AC096910	Rattus no	C 148	17	65.4	177577	3	AC022349	Drosophila
C 76	17.2	66.2	229694	2	AC128965	Rattus no	C 149	17	65.4	179016	3	AC018489	Drosophila
C 77	17.2	66.2	246922	1	AC128068	Rattus no	C 150	17	65.4	187795	2	AC025405	Homo sapi
C 78	17.2	66.2	348014	1	EX640430	Bordetell	C 151	17	65.4	188561	2	AC110739	Mus muscu
C 79	17.2	66.2	349876	6	BX640442	Bordetell	C 152	17	65.4	192001	2	AC068376	Homo sapi
C 80	17.2	66.2	349980	6	AX492782	Sequence	C 153	17	65.4	192372	2	AC110301	Homo sapi
C 81	17.2	66.2	349980	6	AX553949	Sequence	C 154	17	65.4	193314	2	AL603866	Mus muscu
C 82	17	65.4	290	3	AY130267	Microcton	C 155	17	65.4	218800	2	AC133102	Mus muscu
C 83	17	65.4	295	3	AY130268	Microcton	C 156	17	65.4	218800	2	AC094410	Rattus no
C 84	17	65.4	304	3	AY130265	Microcton	C 157	17	65.4	218912	2	AC128733	Rattus no
C 85	17	65.4	315	3	AF428043	Microcton	C 158	17	65.4	224371	2	AC132456	Mus muscu
C 86	17	65.4	329	3	AF428045	Microcton	C 159	17	65.4	229200	2	AC097163	Rattus no
C 87	17	65.4	329	3	AY130263	Microcton	C 160	17	65.4	239923	1	SCO93125	Streptomy
C 88	17	65.4	329	3	AY130264	Microcton	C 161	17	65.4	239923	1	AP005042	Streptomy
C 89	17	65.4	329	3	AY130266	Microcton	C 162	17	65.4	300029	8	AE017057	Oryza sat
C 90	17	65.4	334	3	AF428044	Microcton	C 163	17	65.4	300029	8	AE017057	Oryza sat
C 91	17	65.4	443	1	AF421322	Unculture	C 164	17	65.4	300469	3	AE003503	Drosophila
C 92	17	65.4	457	3	ACABETU	Acanthamoeb	C 165	17	65.4	300550	1	AP005030	Streptomy
C 93	17	65.4	457	3	ACABETU	Acanthamoeb	C 166	17	65.4	302007	1	SCO93132	Streptomy
C 94	17	65.4	457	3	ACABETU	Acanthamoeb	C 167	17	65.4	303050	1	BX321860	Nitrosomo
C 95	17	65.4	461	8	TTR555617	Trametes	C 168	17	65.4	310029	1	AE016874	Pseudomon
C 96	17	65.4	612	8	RHTBETU	Rhodotorula	C 169	17	65.4	310967	1	AE016869	Pseudomon
C 97	17	65.4	706	11	PM12B3G	Penicilliu	C 170	17	65.4	340050	1	MLEBPTN1	Mycoacte
C 98	17	65.4	709	11	PMGEGG	Penicilliu	C 171	17	65.4	347137	1	EX640448	Bordetell
C 99	17	65.4	723	6	AB6149	Sequence 80	C 172	17	65.4	349305	1	BX640433	Bordetell
C 100	17	65.4	725	6	AR155642	Sequence	C 173	17	65.4	349346	1	EX640418	Bordetell
C 101	17	65.4	725	6	E66167	Genome DNA	C 174	17	65.4	349380	6	AX492785	Sequence
C 102	17	65.4	879	6	BD251000	Nucleic a	C 175	17	65.4	349980	6	AX553952	Sequence
C 103	17	65.4	879	6	AR409440	Sequence	C 176	16.8	64.6	1785	6	AX654077	Sequence
C 104	17	65.4	1228	8	AY325811	Alternari	C 177	16.8	64.6	2303	8	AK072823	Oryza sat
C 105	17	65.4	1233	3	AF205379	Leishmani	C 178	16.8	64.6	3141	6	AE01473	DNA encodin
C 106	17	65.4	1722	1	RHMNIFB	R.melliloti	C 179	16.8	64.6	5273	6	B01449	Genomic DNA
C 107	17	65.4	1942	3	AY069706	Oryza sat	C 180	16.8	64.6	5274	6	PSEBDO	Pseudomonas
C 108	17	65.4	1951	8	AK067034	Oryza sat	C 181	16.8	64.6	5274	6	B01338	Genomic DNA
C 109	17	65.4	2198	1	STMREDD	Sequence	C 182	16.8	64.6	6716	1	PSETQDCLC	Pseudomon
C 110	17	65.4	2913	3	CEU132700	Caenorhab	C 183	16.8	64.6	11156	1	AF180147	Pseudomon
C 111	17	65.4	3546	3	CEU132701	Caenorhab	C 184	16.8	64.6	15167	1	PFU18245	Pseudomonas
C 112	17	65.4	5913	1	RER292535	Rhodococc	C 185	16.8	64.6	7207	2	AC074067	Homo sapi
C 113	17	65.4	6027	2	AC013143	Drosophila	C 186	16.8	64.6	100800	8	AC105260	Oryza sat
C 114	17	65.4	9971	1	AE004495	Pseudomon	C 187	16.8	64.6	134247	8	AP004158	Oryza sat
C 115	17	65.4	9979	1	AE007234	Sinorhizo	C 188	16.8	64.6	142854	8	AC136216	Oryza sat
C 116	17	65.4	11610	1	AE004487	Pseudomon	C 189	16.8	64.6	152813	2	AP005860	Oryza sat
C 117	17	65.4	11632	1	AE014727	Bifidobac	C 190	16.8	64.6	152813	2	AC11827	Apis mell
C 118	17	65.4	12016	1	AE011702	Xanthomon	C 191	16.8	64.6	187916	2	AP003458	Oryza sat
C 119	17	65.4	12839	1	AE004617	Pseudomon	C 192	16.8	64.6	192685	1	AE011863	Deinococc
C 120	17	65.4	13187	1	AE005701	Caulobact	C 193	16.8	64.6	198677	1	AE011863	Deinococc
C 121	17	65.4	37566	1	MSGB971CS	L78821 Mycobacteri	C 194	16.8	64.6	203868	9	AP002381	Homo sapi
C 122	17	65.4	42432	1	AF083072	Cenarchae	C 195	16.8	64.6	261160	2	AC140208	Mus muscu
C 123	17	65.4	42432	6	BD250982	Nucleic a	C 196	16.8	64.6	262129	2	AC057899	Rattus no
C 124	17	65.4	42432	6	AR409422	Sequence	C 197	16.8	64.6	346837	1	AP002995	Mesothizo
C 125	17	65.4	70206	2	AC126804	Continuation (4 of	C 198	16.8	64.6	349315	1	EX572593	Rhodopsin
C 126	17	65.4	93407	8	AP006369	Lotus cor	C 199	16.6	63.8	447	8	CPA245656	Cyanophor
C 127	17	65.4	93483	6	AC695974	Sequence	C 200	16.6	63.8	762	6	AR386170	Sequence
C 128	17	65.4	103146	9	HS187N21	Z98036 Human DNA s	C 201	16.6	63.8	889	3	PHTRNAGE	Phycophthor
C 129	17	65.4	110000	2	AC138524	Homo sapi	C 202	16.6	63.8	965	3	AF468039	Oikopleur
C 130	17	65.4	110960	3	CEY39A1A	Caenorhab	C 203	16.6	63.8	1053	8	AK073613	Oryza sat
C 131	17	65.4	112031	9	AL445487	Human DNA	C 204	16.6	63.8	1053	1	AF326408	Pseudomon
C 132	17	65.4	120330	8	AP004261	Oryza sat	C 205	16.6	63.8	1056	3	AF063422	Rubiflex c
C 133	17	65.4	129757	8	ATP2K35	Ala32956 Arabidops	C 206	16.6	63.8	1235	8	FSOCHYC	FSOCHYC
C 134	17	65.4	136197	9	AC105925	Homo sapi	C 207	16.6	63.8	1305	6	AR389262	Sequence
C 135	17	65.4	137560	6	AX754989	Mouse DNA	C 208	16.6	63.8	1650	3	AY118740	Drosophila
C 136	17	65.4	140219	10	AL603714	Mouse DNA	C 209	16.6	63.8	1827	8	CRU40465	Chlamydom
C 137	17	65.4	141293	2	AP004337	Oryza sat	C 210	16.6	63.8	2632	14	AF224760	Cucurbit
C 138	17	65.4	146199	2	AP005107	Oryza sat	C 211	16.6	63.8	2632	14	AF256200	Cucurbit

C 212	16.6	63.8	3077	6	AX598870	Sequence
C 213	16.6	63.8	4218	6	AX647425	Sequence
C 214	16.6	63.8	4986	8	CRU42379	Sequence
C 215	16.6	63.8	6077	6	AX346757	Sequence
C 216	16.6	63.8	6077	6	AX767474	Sequence
C 217	16.6	63.8	6077	6	AX822332	Sequence
C 218	16.6	63.8	6077	6	AX825972	Sequence
C 219	16.6	63.8	7543	9	HS24249601	Sequence
C 220	16.6	63.8	8895	1	BAU51683	Sequence
C 221	16.6	63.8	10029	1	AF263011S2	Sequence
C 222	16.6	63.8	10619	1	AE005683	Sequence
C 223	16.6	63.8	10737	1	AE005269	Sequence
C 224	16.6	63.8	10846	1	AE005144	Sequence
C 225	16.6	63.8	10900	1	AE009524	Sequence
C 226	16.6	63.8	10978	1	AE004795	Sequence
C 227	16.6	63.8	11272	1	AE000189	Sequence
C 228	16.6	63.8	11866	1	AE012109	Sequence
C 229	16.6	63.8	12743	1	AE014416	Sequence
C 230	16.6	63.8	12787	1	AE005170	Sequence
C 231	16.6	63.8	13009	1	AE014811	Sequence
C 232	16.6	63.8	13411	1	AE002035	Sequence
C 233	16.6	63.8	13796	1	D90725	Sequence
C 234	16.6	63.8	14074	1	AE007198	Sequence
C 235	16.6	63.8	14303	1	AE012910	Sequence
C 236	16.6	63.8	15085	1	AE002038	Sequence
C 237	16.6	63.8	17685	1	AE005064	Sequence
C 238	16.6	63.8	27425	6	BD250022	Sequence
C 239	16.6	63.8	31638	1	RMXPNGS	Sequence
C 240	16.6	63.8	34769	6	AX251534	Sequence
C 241	16.6	63.8	44266	1	AB947548	Sequence
C 242	16.6	63.8	61971	8	EX842628	Sequence
C 243	16.6	63.8	67832	6	AX595692	Sequence
C 244	16.6	63.8	86294	5	AC095019	Sequence
C 245	16.6	63.8	88509	2	AC138753	Sequence
C 246	16.6	63.8	93609	2	AC141661	Sequence
C 247	16.6	63.8	97811	2	AC019980	Sequence
C 248	16.6	63.8	98576	2	AC020220	Sequence
C 249	16.6	63.8	105696	10	AL713958	Sequence
C 250	16.6	63.8	110000	2	LMFLCHR36_12	Sequence
C 251	16.6	63.8	111354	8	AP003806	Sequence
C 252	16.6	63.8	122834	2	AC145181	Sequence
C 253	16.6	63.8	124366	8	AC131374	Sequence
C 254	16.6	63.8	131371	9	AC091849	Sequence
C 255	16.6	63.8	131883	8	AC091849	Sequence
C 256	16.6	63.8	142667	10	AC079218	Sequence
C 257	16.6	63.8	150465	8	AC091749	Sequence
C 258	16.6	63.8	152040	2	AP003609	Sequence
C 259	16.6	63.8	152884	9	AC138625	Sequence
C 260	16.6	63.8	155544	8	AP005799	Sequence
C 261	16.6	63.8	156638	6	AX492788	Sequence
C 262	16.6	63.8	156646	6	AX553955	Sequence
C 263	16.6	63.8	158420	9	AC137788	Sequence
C 264	16.6	63.8	161970	8	AP003221	Sequence
C 265	16.6	63.8	163687	3	AC008200	Sequence
C 266	16.6	63.8	165497	8	AP003232	Sequence
C 267	16.6	63.8	166254	2	BX546502	Sequence
C 268	16.6	63.8	173373	3	AC008349	Sequence
C 269	16.6	63.8	176589	3	HS172N19	Sequence
C 270	16.6	63.8	177127	3	AC008199	Sequence
C 271	16.6	63.8	177323	2	AC091933	Sequence
C 272	16.6	63.8	177735	3	AC005714	Sequence
C 273	16.6	63.8	178178	10	AL627312	Sequence
C 274	16.6	63.8	178963	10	AC091503	Sequence
C 275	16.6	63.8	183060	9	AC018641	Sequence
C 276	16.6	63.8	185921	2	AC132645	Sequence
C 277	16.6	63.8	189464	2	AC112194	Sequence
C 278	16.6	63.8	191078	2	AC022505	Sequence
C 279	16.6	63.8	191346	1	AF016485	Sequence
C 280	16.6	63.8	191346	1	AF016485	Sequence
C 281	16.6	63.8	191795	9	AC026748	Sequence
C 282	16.6	63.8	196050	1	AL646058	Sequence
C 283	16.6	63.8	197294	9	AC084125	Sequence
C 284	16.6	63.8	199035	1	AB088420	Sequence

285	16.6	63.8	200062	4	AC091623	Sus scrofa
C 286	16.6	63.8	200441	2	AC022130	Homo sapiens
C 287	16.6	63.8	205486	8	AC074218	Mus musculus
C 288	16.6	63.8	215241	8	AF459639	Triticum aestivum
C 289	16.6	63.8	229887	2	AC073297	Mus musculus
C 290	16.6	63.8	232080	1	AF000003	Pyrococcus furiosus
C 291	16.6	63.8	233000	1	AP000003	Drosophila melanogaster
C 292	16.6	63.8	233356	3	AE003741	Rattus norvegicus
C 293	16.6	63.8	243829	2	AC094446	Rattus norvegicus
C 294	16.6	63.8	257621	2	AC094703	Rattus norvegicus
C 295	16.6	63.8	260050	1	SME591782	Sinorhizobium meliloti
C 296	16.6	63.8	280265	2	AC132159	Rattus norvegicus
C 297	16.6	63.8	293050	1	SC0939116	Streptococcus pneumoniae
C 298	16.6	63.8	297816	1	AP002553	Escherichia coli
C 299	16.6	63.8	300000	1	SME591784	Sinorhizobium meliloti
C 300	16.6	63.8	300029	8	AE017076	Oryza sativa

ALIGNMENTS

RESULT 1	AB149059/c	AB149059	401 bp	DNA	linear	PAT 08-AUG-2001
LOCUS	Sequence 75 from patent US 6228371.					
DEFINITION	Sequence 75 from patent US 6228371.					
ACCESSION	AB149059					
VERSION	AB149059.1	GI:15113650				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 401)					
AUTHORS	Nano, F.E.					
TITLE	Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides					
JOURNAL	Patent: US 6228371-A 75 08-MAY-2001;					
FEATURES	Location/Qualifiers					
source	1..401					
organism	/organism="unknown"					
mol_type	/mol_type="unassigned DNA"					

Query Match	100.0%;	Score 26;	DB 6;	Length 401;
Best Local Similarity	100.0%;	Pred. No. 0.45;		
Matches	26;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

Qy	1	ACAACGGGAGCGCCGAATCTACGAA	26
Db	148	ACAACGGGAGCGCCGAATCTACGAA	123

RESULT 2	AR340431/c	AR340431	401 bp	DNA	linear	PAT 17-AUG-2003
LOCUS	Sequence 75 from patent US 6572865.					
DEFINITION	Sequence 75 from patent US 6572865.					
ACCESSION	AR340431					
VERSION	AR340431.1	GI:33731979				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 401)					
AUTHORS	Nano, F.E.					
TITLE	Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides and methods for using same					
JOURNAL	Patent: US 6572865-A 75 03-JUN-2003;					
FEATURES	Location/Qualifiers					
source	1..401					
organism	/organism="unknown"					
mol_type	/mol_type="genomic DNA"					

Query Match	100.0%;	Score 26;	DB 6;	Length 401;
-------------	---------	-----------	-------	-------------

Best Local Similarity 100.0%; Pred. No. 0.45; Mismatches 0; Indels 0; Gaps 0;
 Matches 26; Conservative 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGAA 26
 Db 148 ACAACGGCGAGGCCGGAATCTACGAA 123

RESULT 3
 AE007182 18121 bp DNA linear BCT 27-APR-2001
 LOCUS Mycobacterium tuberculosis CDC1551, section 268 of 280 of the
 DEFINITION complete genome.
 ACCESSION AE007182 AE000516
 VERSION AE007182.1 GI:13883751
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis CDC1551
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.
 1 (bases 1 to 18121)
 Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
 Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
 Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
 Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H.,
 Gill, J., Mikula, A. and Bishai, W.
 Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains
 Unpublished
 2 (bases 1 to 18121)
 Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
 Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
 Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
 Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H.,
 Gill, J., Mikula, A. and Bishai, W.
 Direct Submission
 Submitted (25-APR-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1..18121
 /organism="Mycobacterium tuberculosis CDC1551"
 /mol_type="genomic DNA"
 /strain="CDC1551"
 /db_xref="taxon:83331"
 /note="clinical strain"
 207..896
 /gene="MT3873"
 207..896
 /gene="MT3873"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48237.1"
 /db_xref="GI:13883752"
 /translation="MSAFDSGSLTIGVIYVARPNWVANVTSMIDAAAGLHPRV
 ALMDVSGNPPGGSWINRLYNLWADYAGSPVRIIGYANAYDFFNWRVRPAGLR
 VIGAGYGNLPGVAHQYTDGSGYSPNLPGQAPFGRCDMNSANGLTPQOFAAACG
 VTTTGGPLMALTDDEQTELTIKAREIWDQIRGPNAGWPQLGQNEQOGLTPEVDAIAV
 IKNDVAAMLAIE"
 complement(910..1854)
 /gene="MT3874"
 complement(910..1854)
 /gene="MT3874"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK48238.1"
 /db_xref="GI:13883753"
 /translation="MPTDNDSWAITESVCATAGVAARAAETESDNLNDPFARI
 FVDAAGDGWSMTYNTLLAGATDLDLRAPIQMIDFMFAARTAFPEYFIATADAG

Qy 184..2343
 Db 184..2343

gene
 CDS
 /gene="MT3875"
 184..2343
 /gene="MT3875"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48239.1"
 /db_xref="GI:13883754"
 /translation="MGSTPRTPOEFAHGHQAAGDLDEIVADYADDSEFVITPAGI
 ARGKGIHQLEFVLLDDIFNALMDLTKIIFEGDILFLEWTANSASVRVDDGVTFVPR
 DGTIWAHTVRYTPHPT"
 complement(2388..2906)
 /gene="MT3876"
 complement(2388..2906)
 /gene="MT3876"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48240.1"
 /db_xref="GI:13883755"
 /translation="MTPSGTTEHNGSVRIHRPOITVCRACKIAQPEGESIVRGAEOK
 VLDGLSLFQRPHPRIQVGLGRELADFGHVAQSSGVFPDGRDSELSVAGARVDGRE
 DGSIVGLVRFQSGDPCSLLQRRHGLISRSAGSSAKCCNLFISCAARNAPPGSI
 RPRQTTPVRSR"
 2996..3127
 /gene="MT3876.1"
 2996..3127
 /gene="MT3876.1"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48241.1"
 /db_xref="GI:13883756"
 /translation="MNETMGCADIGRWYGLVADPVRARGHPSPDVEDRALITWR"
 complement(3114..3689)
 /gene="MT3877"
 complement(3114..3689)
 /gene="MT3877"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48242.1"
 /db_xref="GI:13883757"
 /translation="MLSGIQQNTLMNDPLAHGYVADLLVALAVVLMARTRPE
 LARMLLCTGLVWELFVGLSAWNTPIEWATPLPTVVFLLAHVMDGGLTLM
 GWLLARALTGEAGALGLTVQVLWGQLTALAVLSAILAGTWSYVDDLVNFWVFWFR
 GHPTAAQMTLWLLPLCEALVRLRLTAR"
 complement(3795..3977)
 /gene="MT3878"
 complement(3795..3977)
 /gene="MT3878"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK48243.1"
 /db_xref="GI:13883758"
 /translation="MGSTPWCNPPCQCTLTPEVLELAVALPENPDRTAGAIQRI
 RAQLAGDRIALRGRGS"
 complement(3992..4183)
 /gene="MT3879"
 complement(3992..4183)
 /gene="MT3879"
 /note="identified by Glimmer2; putative"

```

RESULT 4
MSGY126
LOCUS       37164 bp DNA linear BCT 10-DEC-1996
MycoBacterium tuberculosis sequence from clone y126.
AD000012
ACCESSION   AD000012
VERSION     1
KEYWORDS    GI:1717733
SOURCE      Mycobacterium tuberculosis
ORGANISM    Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1 (bases 1 to 37164)
Du, L.
Direct Submission
Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham, MA, USA, 02154 du@eric.com
GSDS:S:1004704
FEATURES             source
     1..37164
         /organism="Mycobacterium tuberculosis"
         /mol_type="genomic DNA"
         /db_xref="taxon:1773"
         /clone="y126.nt"
ORIGIN
Query Match      100.0%; Score 26; DB 1; Length 37164;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ACAACGGCGAGGCCGAATCTACGAA 26
          |||||
Db       13576 ACAACGGCGAGGCCGAATCTACGAA 13601
          |||||
RESULT 5
BX842584
LOCUS       244800 bp DNA circular BCT 21-NOV-2003
MycoBacterium tuberculosis H37Rv complete genome; segment 13/13.
ACCESSION   BX842584
VERSION     1
KEYWORDS    AL021426 AL022076 AL022120 AL022121 AL123456 Z80343 Z83864
SOURCE      Mycobacterium tuberculosis H37Rv
ORGANISM    Mycobacterium tuberculosis H37Rv
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
Tekai, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S., and
Squares, S., Squares, R., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature 393 (5685), 537-544 (1998)
98295987
PUBMED     9634230
REFERENCE   2
AUTHORS    Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.
Re-annotation of the genome sequence of Mycobacterium tuberculosis
H37Rv
Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
JOURNAL    Microbiology
MEDLINE    12368430
PUBMED     12368430
REFERENCE   3
AUTHORS    Parkhill, J.
Re-annotation of the genome sequence of Mycobacterium tuberculosis
H37Rv
Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
JOURNAL    Microbiology
MEDLINE    12368430
PUBMED     12368430

```


/locus_tag="Mb3713c"
/notes="Mb3713c, -, len: 154 aa. Equivalent to Rv3688c, len: 154 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 154 aa overlap). Hypothetical protein, similar to other bacterial hypothetical proteins e.g. Q9X934[SCHE66.02c from Streptomyces coelicolor (154 aa), FASTA scores: opt: 425, E(): 3.4e-19, (46.1% identity in 154 aa overlap); Q9WZF4[TM0690 from Thermotoga maritima (149 aa), FASTA scores: opt: 326, E(): 3.4e-13, (40.4% identity in 151 aa overlap); Q9PHU3[CU0573 from Campylobacter jejuni (147 aa), FASTA scores: opt: 290, E(): 5.1e-11, (36.4% identity in 151 aa overlap); etc. Also some similarity to upstream
Q9654[RV3686c|MTV035_034c CONSERVED HYPOTHETICAL PROTEIN from Mycobacterium tuberculosis."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD95899.1"
/db_xref="GI:31620457"
/translation="MSELKSLRDLTQAMKTDQKLRATIRMLLAAQTQTEVSGKQA
RELSDDVIVKILARESRRGEAAIYTCNGSELAATPHAEARIIDEVLTPTLTGEG
ADVADTAIEVAEELGHRPSKMQGLVWKAAVTIAGKAGARLSAAVKERL"
704..2059
/locus_tag="Mb3714"
704..2059
/locus_tag="Mb3714"
/notes="Mb3714, -, len: 451 aa. Equivalent to Rv3689, len: 451 aa, from Mycobacterium tuberculosis strain H37Rv, (99.8% identity in 451 aa overlap). Probable conserved transmembrane protein, with proline rich N-terminus, similar to Q9KW6[SCB33.17 PUTATIVE INTEGRAL MEMBRANE PROTEIN from Streptomyces coelicolor (462 aa) FASTA scores: opt: 730, E(): 2.7e-21, (38.1% identity in 412 aa overlap)."
/codon_start=1
/transl_table=11
/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN"
/protein_id="CAD95900.1"
/db_xref="GI:31620458"
/translation="MHKRAYAPQPKPDTETYEKTRDRQDGHERRQLRPVMSLPG
PGVPEVPVAPGAPVAPGVPYAPATPGVPGYGCAPSYGPPGPGPLGYPAPPG
CGPPGVPPLGSGPELAPGAKPGIIEPLRLTDSIENGAVGYIRANPKATILGTAM
VWVLIQISIALVGPMTAPGAVITVTEGEDELTVGVVGWGSFGASGLVSMVLGVLSS
GMLTVGVRAPGPTVTEGAWKVRGELLALFLALLEAAGVAVLGAIVLTLSGVA
AAENAAALGLPELLVGVNSLAYLVLLFAPVLLVLERIPVEATIRTSALVRHG
FWRLGRLTLLVGVVGNAAFPAMFVIGEIVTAVTADSGVTKRLVGATLSAIGVT
ICQIVTAPFSAGVVLVLTDRRIAEAFDLVQTGLEAGPAGPAPVESTDLNLWLRP
F"
2086..2739
/locus_tag="Mb3715"
2086..2739
/locus_tag="Mb3715"
/notes="Mb3715, -, len: 217 aa. Equivalent to Rv3690, len: 217 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 217 aa overlap). Probable conserved membrane protein, similar to Q9KW5[SCB33.18 PUTATIVE INTEGRAL MEMBRANE PROTEIN from Streptomyces coelicolor (231 aa), FASTA scores: opt: 419, E(): 1.5e-19, (36.0% identity in 211 aa overlap). Equivalent to AA48159 from Mycobacterium tuberculosis strain CDC1551 (233 aa) but shorter 16 aa."
/codon_start=1
/transl_table=11
/product="PROBABLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAD95901.1"
/db_xref="GI:31620459"
/translation="MPSIDIDREAHQAQARELDKPIYFDKSLTKELTDWIDRLQRYRI
LEKSGSIPEGWFTIVTLLILMLAVTAQCIARNTWRNGGDIYKFDAGQLTAQHR
STAEGYAAEGWAAAIARHRLQAVARELETKMLNPAAGTANELASDAGDFPILAGE
LTQATNFNDVYIGRPGTQAGYMIADLDDHLSRSPAVSVSQHPAFVDFSAQVR"
2865..3866
/locus_tag="Mb3716"

CDS

2865...3866
 /locus tag="Mb3716"
 /note="Mb3716", len: 333 aa. Equivalent to Rv3691, len: 333 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 333 aa overlap). Conserved hypothetical protein, similar to Q9KYM4|SCE33.19 PUTATIVE SECRETED PROTEIN from Streptomyces coelicolor (387 aa)
 PASTA scores: opt: 481, E(): 6e-23, (36.6% identity in 358 aa overlap). Equivalent to AAK48160 from Mycobacterium tuberculosis strain CDC1551 (381 aa) but shorter 48 aa.
 /codon_start=1
 /transl_table=11
 /product="CONSERVED HYPOTHETICAL PROTEIN"
 /protein_id="CAD95902.1"
 /db_xref="GI:31620460"
 /translation="MAPASTSSSTGGHALATLGNHGVVVVADSIADVEAARPDILL
 LVAOTYLDNALDLRAKAPGDLVAPTSRTTALTPOLRILAAAPFNSQPNCTLR
 EANPAGSVQPGSPDYCATGDLVLTSCYGGALYRFAEGRTIIVGSSHPMTNGLLP
 AGNALANLGNRPRLVITAPDIEGEMSPSLDLIPENVHTIWLQWLVLLVA
 LMKRRIGPLVAELPVIIVRASETVEGRGLYRSRRARDRAADALTATLQRLRPLG
 VGAGAPAVVVITIAQRSKADPPFPVAYHLFGPAPATDNDLLQLARALDDIERQVHS"
 3863...4939
 /gene="moxr2"
 /locus tag="Mb3717"
 /note="Mb3717", len: 358 aa. Equivalent to Rv3692, len: 358 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 358 aa overlap). Probable moxr2, methanol dehydrogenase regulatory protein, highly similar (generally longer at N-terminus) to Q9KIW3|SCE33.20 PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (329 aa), PASTA scores: opt: 1523, E(): 4.2e-74, (70.9% identity in 330 aa overlap); Q92538|SC9B2.21c PUTATIVE REGULATORY PROTEIN from Streptomyces coelicolor (332 aa) PASTA scores: opt: 1008, E(): 1.1e-46, (50.8% identity in 313 aa overlap); Q9UZ67|MOXR-3|PAB0848 METHANOL DEHYDROGENASE REGULATORY PROTEIN from Pyrococcus abyssi (314 aa), PASTA scores: opt: 989, E(): 1.1e-45, (50.65% identity in 302 aa overlap); Q9AAN1|CC0566 MOXR PROTEIN from Caulobacter crescentus (323 aa), PASTA scores: opt: 988, E(): 1.3e-45, (52.3% identity in 306 aa overlap); etc. Also similar to O53170|MTV007.26|MOXR|Rv1479 from Mycobacterium tuberculosis (377 aa); and O07392|AF002133.6|MOXR from Mycobacterium avium (309 aa). Also high similarity with several hypothetical bacterial proteins.
 /codon_start=1
 /transl_table=11
 /product="PUTATIVE METHANOL DEHYDROGENASE TRANSCRIPTIONAL REGULATORY PROTEIN MOXR2"
 /protein_id="CAD95903.1"
 /db_xref="GI:31620461"
 /translation="MTQSANPQAPPTQTGAEPLPGPAGAPTAAPSGPHPHRAE
 AESARDALALRAEVAKAVGQGVISGLVIALLCRHVLEGVCAKTLIVRMSA
 ALQLEKRVQFTDPLMGVDVTSGLVDAFTAFVFPFVFNLIADINTPKTKO
 AALLEAMEERQGEKPLPNPFIIVATNPQIEYGVQLPEAQDRFLKLVNLT
 PARDEIATDRHAHGFDRDLASINPVAAGPAELAGREAVRHVLANEVLGYIVDVI
 GATESPALQGVSPGATALLGTARSNAWLSGSDVVTDDVKANARFILRHVMRLP
 EASLEGATPDGVLDGLIASFVVR"
 5073...6395
 /locus tag="Mb3718"
 /note="Mb3718", len: 440 aa. Equivalent to Rv3693, len: 440 aa (alternative start at 41910), from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 440 aa overlap). Possible conserved membrane protein, similar to Q9KIW2|SCE33.21 PUTATIVE LIPOPROTEIN from Streptomyces coelicolor (436 aa), PASTA scores: opt: 875, E(): 3.3e-46, (56.25% identity in 448 aa overlap); Q9AAN0|CC0567 HYPOTHETICAL PROTEIN from Caulobacter crescentus (437 aa),

gene

CDS

gene
CDSFEATURES
sourceTITLE
JOURNALREFERENCE
AUTHORSTITLE
JOURNALREFERENCE
AUTHORSREFERENCE
AUTHORSREFERENCE
AUTHORSREFERENCE
AUTHORSREFERENCE
AUTHORSREFERENCE
AUTHORSREFERENCE
AUTHORSREFERENCE
AUTHORS

PASTA scores: opt: 355, E(): 2.3e-14, (30.9% identity in 450 aa overlap); P73233|SLR2013 HYPOTHETICAL 48.5 KDA PROTEIN from Synechocystis sp. strain PCC 6803 (435 aa)
 PASTA scores: opt: 340, E(): 1.9e-13, (29.7% identity in 438 aa overlap); etc. Equivalent to AAK48162 from Mycobacterium tuberculosis strain CDC1551 (475 aa) but shorter 35 aa. Also similar to other hypothetical proteins from Mycobacterium tuberculosis; MTV014_7; MTV007_27; and MTCV71_36 M.
 /codon_start=1
 /transl_table=11
 /product="POSSIBLE CONSERVED MEMBRANE PROTEIN"
 Query Match 100.0%; Score 26; DB 1; Length 278492;
 Best Local Similarity 100.0%; Pred. No. 0.43;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACAACGGCGAGCCCGCAATCTACGAA 26
 |||||
 DB 96222 ACAACGGCGAGCCCGCAATCTACGAA 96247
 |||||
 RESULT 7
 AE007304/c 12461 bp DNA linear BCT 15-AUG-2001
 LOCUS Sinorhizobium meliloti plasmid pSymb section 110 of 121 of the complete plasmid sequence.
 ACCESSION AE007304 AE006469
 VERSION AE007304.1 GI:14524346
 KEYWORDS
 SOURCE Sinorhizobium meliloti (Rhizobium meliloti)
 ORGANISM Sinorhizobium meliloti
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 REFERENCE 1 (bases 1 to 12461)
 AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P., Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J., Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L., Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R., Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
 Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymb megaplasmid
 Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
 MEDLINE 21396509
 PUBMED 11481432
 REFERENCE 2 (bases 1 to 12461)
 AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P., Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J., Gurjal,M., Hong,A., Huizar,L., Hyman,R.W., Kahn,D., Kahn,M.L., Kaiman,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R., Wells,D.H., Yeh,K.-C., Davis,R.W., Federspiel,N.A. and Long,S.R.
 Direct Submission
 Submitted (29-MAR-2001) Biological Sciences, Stanford University, 371 Serra Mall, Stanford, CA 94305, USA
 Location/Qualifiers
 1. 12461
 /organism="Sinorhizobium meliloti"
 /mol_type="genomic DNA"
 /strain="1021"
 /db_xref="taxon:382"
 /plasmid="pSymb"
 complement(134).520)
 /gene="Sma2191"
 complement(134).520)
 /gene="Sma2191"
 /function="Miscellaneous; Hypothetical/Global homology"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK65841.1"
 /db_xref="GI:14524347"
 /translation="MRYGMRVEILGQRRRRWGDAKKLDIVMSIGLDGATVTEVAHRH
 DVSRQIYVNRHLEKKKGLLPSTADIVFLPVDVNAHMPVREDVARYVSAMELRIN

```

DGRSLRFDSDGVDPAVLTIRAVEAA"
/translation="MSAVKANRYQTAVAAATWAVAVGVLTGAVHAQTAADILPOLYR
complement (575. .1888)
/gene="Sma2193"
CDS
complement (575. .1888)
/gene="Sma2193"
/function="Small molecule metabolism"
/notes="glimmer prediction; similar to agaE [Agrobacterium
tumefaciens], gb|AA88469. Weak similarity to pyridine
nucleotide-disulphide oxidoreductase family"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase"
/protein_id="AAK65842.1"
/db_xref="GI:14524348"
/translation="MTRQGIQPSQKWDLIIVGGIGVGSTAFYAARLGLRVLIIVERD
TPSGQSGNIGFVRQOQDFRELPLAIALRLMNDLEDLGRKVGWLRGNNVLAAN
DRTAQAQWQAKDFGLDTLLTSAQTRKPLPLSDKASLCGAMFTASDGRAPGR
TTPAFAQWQKSVILGAVHTQELQAGNVHGVNDKLYRADRVLCASCTGSGKL
LRGVGNLQBRIRAVARTLPARGTLTSCISLPTGLRQDEKGFIESVAGGYDV
RFDUSWLYHYRTRRANPDARVNTFGILRPMSSRIASPIADIAPSEGARPAHYR
VQAQDELQFQFLAEALAEAVAGVIDALPDVPMVGHVYERPGLLIATGFGSHGF
GLGPWAGKIMAGLAAGHSADVDSGLSPSRFHSLSVTARLASPAG"
complement (1889. .2653)
/gene="Sma2195"
CDS
complement (1889. .2653)
/gene="Sma2195"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to amino acid and opine
transporters, e.g., Agrobacterium tumefaciens P35117.
ABC transporter family"
/codon_start=1
/transl_table=11
/product="probable ABC transporter, ATP-binding protein"
/protein_id="AAK65843.1"
/db_xref="GI:14524349"
/translation="MSHPITVFDRIYVDRVSKSYGSMVTINDINLEIATGTSVLSIGPSGSK
STLLRCVNHLEIRERSITVDGELIGYRRVGNLAHELPEKLVARQAGIGMVFQFNL
FGHKTALNVEIAPVHVRGLARDAEQATALLERVLADQWQVPRMLSGQOQORVA
IARLAKPKVILPDPTSLADPELAGEVLDVMSIAEGLTVMVVTHEMSFAREVCN
RIAPMAGEIVECAPSEFDPKPAFORTSRFLSKVH"
complement (2655. .3569)
/gene="Sma2197"
CDS
complement (2655. .3569)
/gene="Sma2197"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to permease AtrB from
Streptomyces coelicolor, AAF25532;
binding-protein-dependent transport system"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, permease"
/protein_id="AAK65844.1"
/db_xref="GI:14524350"
/translation="MSNDLTIAPTASDPRLTYVPASTWTKRFAATLLILLVLYCAT
MIASNPFGVDVADYLDSDRLWGLSLTVMVTWIGVLTGTFPAINMADNVVI
STVANAYWFRFTPLVQLIPWNLGALFPQLSVGVPTSLVSTPTLSPVTA
VLGLNNEGMYSEIIRSGMSVDPQORQAASGLMGNTGKTLWRILLPQAMPYIIPPT
GNOTIGMLKTSLSLVISLADLYSAQTIYSRNFQIPLLIIVACIWLAAATLISAVQ
VRIERPARSORANITQPERLFRQAR"
complement (3633. .4523)
/gene="Sma2199"
CDS
complement (3633. .4523)
/gene="Sma2199"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; putative bacterial
extracellular solute-binding protein, similar to
Streptomyces coelicolor AtrA, AAF25531"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, periplasmic
solute-binding protein"
/protein_id="AAK65845.1"
/db_xref="GI:14524351"
/translation="MSAVKANRYQTAVAAATWAVAVGVLTGAVHAQTAADILPOLYR
DAGVILVTDAKYPPFQSVNDAGEIVGFVDLWNAIADRNLNVKVDVTSVFSDSLPGV
QGRWDIAMEGITDNAERQKVFYDGYTTSSAYLEQSGSEIKDHLGCGKGSQAQ
SGTEWGMIAKEIGACVAAKQKPSYBFGTSEATILLSYSGRSDFLVLSAALAGEI
QKVAPHVQVIVPMQILPRMPSGIAFRKDETDLDGALLALKLKEIRANGDYEKIVAKTV
SPMAVEHEPGINLATVPQTK"
complement (4635. .5015)
/gene="Sma2201"
CDS
complement (4635. .5015)
/gene="Sma2201"
/function="Miscellaneous; Unknown"
/notes="glimmer prediction; some similarity to putative
translation initiation inhibitors. similar to SMA2289 of
S. melliloti pSymba."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK65846.1"
/db_xref="GI:14524352"
/translation="WKXRIFAGVPSSEEMAGYAKAVIVGTTVVSGTTGRDPTDGNFPD
NAAQARNALASIDKALRQAGSLANAVSRVVVDQACADVTAVLGEVFRDIRPTS
TMLICEIPAPGAKVEIEITABLEC"
complement (5106. .5900)
/gene="Sma2203"
CDS
complement (5106. .5900)
/gene="Sma2203"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to transport protein
PotC [Actinobacillus actinomycetemcomitans]: gb|AAC27497"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, permease"
/protein_id="AAK65847.1"
/db_xref="GI:14524353"
/translation="MTRWDISSIALRLVTFSSMAFLPIPIAVTLIVAVNPREFVLPPN
GFTLDWFAAASSXTFLRGMGVSLVGLAALIANALAPAAIALARNRNFPGKAINL
LMTPLTLTILSALYIFRVYGGSLGVLLVGHAIHMPYAVRILTSALNFDP
SVEEAAARVAGRIITFVTLPTVITGLISLTLCSLWNPDPPIVSFLAPPWTPL
PVELYVTKFOYDAVAALASSLILLSAVAMVWIDRLAGLRRLRS"
complement (5897. .6766)
/gene="Sma2205"
CDS
complement (5897. .6766)
/gene="Sma2205"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to PotB permease for
spermidine and putrescine [E. coli]: P23860"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, permease"
/protein_id="AAK65848.1"
/db_xref="GI:14524354"
/translation="MPENTRILPLSPALLALLFVAPMVWVVRTSFNLAEGAYI
VEAMTLENVTRFLTEPVTWNTLWFSVRIALATAISVICAPVALYIAKTSGLQRY
LMALTMALILGLVTLVYGVTVIPRGGGLMSLALRVYDQPVRYMDIKGVVILLV
YIGTPIVLSLDSIERINFLVEARNVGNRTAFWKIVPELSTPGLYAGLVVRS
LNFAFAPVLEMGDSNTOMIGLVIYREALNNDLPFASALSIVWVSNALLILGMSAL
AGRLILSLRAKR"
complement (6759. .7844)
/gene="Sma2207"
CDS
complement (6759. .7844)
/gene="Sma2207"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction; similar to putative
ABC-transporter AtrA required for attachment to host
cells and for virulence [Agrobacterium tumefaciens]:
AAB67294"
/codon_start=1
/transl_table=11
/product="putative ABC transporter, ATP-binding protein"
/protein_id="AAK65849.1"
/db_xref="GI:14524355"
/translation="MSNDSIQVEGITKRYGAMTAVDNVDFVQSGQFVSLGLPSCGCK
TTTLRIAGFVDPGGLLRVRGRDVTLLPPEKRDILGFVFQNYALWPHMTVAENVAFGL
```

KLRKKNRFRDKVAELTTTGLSGYENRRLPQLSGGQORVALARALALEPQVLMD
EPLNLRALRVTRRELKELQARKMTLYVTHDQELASMGNVVNNKGVQVQA
APFLYENPFGFVADFGITNFGGIVTSAEGGVGVKLGSGQMLKAVTKQPVALGT
KVRALIRPVRTLTAAADGNILTGQVLAEGYFGALLRYSLRDLSDILRAEVHNF
SFIESGSRVWVCVSPHEHLRLIPDERNA"

Query Match 75.4%; Score 19.6; DB 1; Length 12461;

Best Local Similarity 84.6%; Pred. No. 3.6e+02;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGGAATCTACGAA 26
|||||
DB 10910 AGAACGGCGTGGCCGGATCTACGAA 10885
|||||

RESULT 8

MSGB27CS/3 38793 bp DNA linear BCT 17-DEC-2001

LOCUS Mycobacterium leprae cosmid B27 DNA sequence.

DEFINITION L78817

ACCESSION L78817.1 GI:137777

VERSION

KEYWORDS

SOURCE

ORGANISM

Myco

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium.

1 (sites)

Eiglmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.

Use of an ordered cosmid library to deduce the genomic organization

of Mycobacterium leprae

Mol. Microbiol. 7 (2), 197-206 (1993)

93188700

8446027

2 (bases 1 to 38793)

Smith, D.R., Richter, P., Rubenfield, M., Rice, P.W., Butler, C.,

Lee, H.M., Kirst, S., Gundersen, K., Abendschan, K., Xu, Q., Chung, M.,

DeLoughery, C., Aldredge, T., Maher, J., Lundstrom, R., Tullig, C.,

Falls, K., Imrich, J., Torrey, D., Engelstein, M., Breton, G., Madan, D.,

Nietupski, R., Seitz, B., Mao, J.I. et al.

Multiplex sequencing of 1.5 Mb of the Mycobacterium leprae genome

Genome Res. 7 (8), 802-813 (1997)

97413161

9267804

Original source text: Mycobacterium leprae (clone: cosmid B27)

(tissue library: Loris 6) DNA.

This sequence was generated by the Genome Sequencing Center at

Genome Therapeutics Corporation (Collaborative Research Division),

100

Beaver St., Walham, MA, 02154. Please contact Doug Smith

(smith@eric.com) for further information. The sequence represents

the

insert of a Loris 6 cosmid clone from a mapped set of clones

constructed from M. leprae genomic DNA isolated from armadillo

liver

[3]. The sequence may not represent the entire cloned insert of

the

cosmid if an overlapping region was previously sequenced from

another

clone. Coding sequences larger than 60 amino acids were predicted

on

the basis of codon usage and homology information. An attempt was

made

to locate the most probable start site based on codon usage,

homology,

the presence of a Shine-Dalgarno sequence, or overlapping orf that

suggested translational coupling. It is possible that the actual

start

site differs from the one selected.

Location/Qualifiers

1..38793

/organism="Mycobacterium leprae"

/mol_type="genomic DNA"

/specific_host="Dasypus novemcinctus"

source

FEATURES

/db_xref="taxon:1769"

/clone="cosmid B27"

/tissue_lib="Loris 6"

/note="The liver of the armadillo was used to isolate the

Mycobacterium leprae."

ORIGIN

Query Match 74.6%; Score 19.4; DB 1; Length 38793;

Best Local Similarity 95.2%; Pred. NO. 4.5e+02;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGGAATCTA 22
|||||
DB 17212 CAACGGCGAGCGCCGGAATCTA 17192
|||||

RESULT 9

MLEPRTN4/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Myco

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium.

1 (bases 1 to 348450)

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R.,

Wheeler, P.R., Honore, N., Ganiher, T., Churcher, C., Harris, D.,

Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,

Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A.,

Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C.,

Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A.,

Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,

Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K.,

Taylor, K., Whitehead, S., Woodward, J.R. and Barrell, B.G.

Massive gene decay in the leprosy bacillus

Nature 409 (6823), 1007-1011 (2001)

21128732

11234002

2 (bases 1 to 348450)

Parkhill, J.

Direct Submission

Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium

leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome

Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique

Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,

75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk

Notes:

Details of M. leprae sequencing at the Sanger Centre are available

from http://www.sanger.ac.uk/Projects/M_leprae/ A relational

database containing the M. leprae sequences is available from

<http://genolist.pasteur.fr/Leprona/>.

Location/Qualifiers

1..348450

/organism="Mycobacterium leprae"

/mol_type="genomic DNA"

/strain="TN"

/db_xref="taxon:1769"

551..1786

/gene="ML0842"

551..1786

/note="Similar to Mycobacterium tuberculosis hypothetical

44.6 kDa protein Rv1464 or MT0007.11 TR:O53155

(EMBL:AL021184) (417 aa) fasta scores: E(): 0, 43.6% id in

408 aa and to many other bacterial nifs-homologues, e.g.

Bacillus subtilis YurW protein yurW TR:O32164

(EMBL:Z99120) (406 aa) fasta scores: E(): 0, 46.7% id in

405 aa. Previously sequenced as TR:O32975 (EMBL:Z98741).

Contains pfam match to entry PF00266 aminotran_5,

Aminotransferases class-V.


```

Similar to ML0117, ML0596 and ML1708"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC31223.1"
/db_xref="GI:13092923"
/db_xref="GOA:O32975"
/db_xref="SWISS-PROT:O32975"
/translation="WDVFDIOARADPFILQETVNGKPLIWFNDNAATQKPOVIDR
LSYFAHNSNTHRAHEALAAATDAIEARETARFAGAKAQSLIFVGTTEAINL
VAYAGGLQPGDEVIITLHSHANIWVQLLSQTGAILKAVPDVDAAGLLMSBEF
DLIGRTKLVATQVNSALGTQGEKIVELGHRYGARVLIDGAQSIPLHPIFNSELG
ADFFVPSGHKIYGPYIGVYCEVLTMPWQGGNMIVDVTLERSLYOQPNKFE
AGTNADAGLEALGVYRVGVORIASHECALDYATPRLADIPGVRLVGTATEKA
SVLSPVLAHEPLEVGLKALNAEGIAVRAGHCAQPVLRRLGLEATVPSFAPYNYEY
IDVFINVRRIRAEAGTINIG"
misc_feature
653..1738
/gene="ML0842"
/notes="Pfam match to entry PF00266 aminotran 5,
Aminotransferases class-V, score 23.70, E-value 1.7e-08"
1908..2395
/gene="lppp"
/pseudo
/notes="synonym: ML0843"
1908..2395
/gene="lppp"
/notes="Possible pseudogene of M. tuberculosis orthologue
lppp (Best blastx score 294)"
/pseudo
/codon_start=1
/transl_table=11
/product="lipoprotein (pseudogene)"
2549..2555
/notes="possible RBS"
2559..4112
/gene="nark"
/notes="synonym: ML0844"
2559..4112
/gene="nark"
/notes="Similar to Mycobacterium tuberculosis putative
nitrite extrusion protein narK1 or Rv2329c or MTCY3G12.05
TR:P71683 (EMBL:Z79702) (515 aa) fasta scores: E(): 0,
69.3% id in 488 aa and to Escherichia coli nitrite
extrusion protein 2 narU SW:NRU_ECOLI (P37738; P77696)
(462 aa) fasta scores: E(): 0, 36.6% id in 459 aa. There
is a frameshift near the C-terminus relative to the M.
tuberculosis homologue. Previously sequenced as TR:O32974
(EMBL:Z98741). Contains hydrophobic, probable
membrane-spanning regions."
/codon_start=1
/transl_table=11
/product="putative nitrite extrusion protein"
/protein_id="CAC31225.1"
/db_xref="GI:13092924"
/db_xref="SPTREMBL:O32974"
/translation="MDQLVLLQAESLYGPRRSKTVPIDHBSLXSSFFRLRRAC
RISHWPEQAGNKGKTIARLNLSWLVITHLGSVWMLPFWELNPKDVIYGSFA
GDFLATTATLVACLRVPYSLATALFGERNWAIFSVWLPFIATNVLHAPGLP
LWPLACALITGLGGNFAASMTNANAFPHRLKGAAGLACAGNLGVSIVQVGMGL
VIASVGDRLPYIVCGLYVLLIIAGICAMLFMNDIEHHRIGVNTIRPVLVIVSTRDS
WVLALVLAISFGSIFGSFAFGQLETFVAGSQSTAOAALHAELAFIGPTLAAVAR
FWGRLADRLGSRVTLVYVFGAMVFAAGLLGVLIIEGSRVCEIRGVMAVYACFIT
LFTLSLGNVSGYKMLPTTFEACHSHSGINDDECDWRSVIGSVIGFVAEYALGGV
GIDLALRESYNTGGVTAAFWIFMLCYAAAGVLTWKMYVCRPLPGLNLDHEAANAFAAS
VGASRTHRG"
complement(4100..4681)
/gene="ML0845"
complement(4100..4681)
/gene="ML0845"
/notes="Unknown function. Similar to part of some acyl-CoA
oxidases e.g. Arabidopsis thaliana acyl-CoA oxidase cx2
TR:O65201 (EMBL:AF057043) (692 aa) fasta scores: E():
1.5e-06, 35.2% id in 125 aa. Previously sequenced as

```

```

TR:O32973 (EMBL:Z98741) (193 aa) fasta scores: E(): 0,
99.5% id in 193 aa."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC31226.1"
/db_xref="GI:13092925"
/db_xref="SPTREMBL:O32973"
/translation="MPLRPTNAASERVTKETATETILQRMVGARQDSEKKSLPNRGT
QGTIFCEKCECLLSAVARRRQVSKENAFDTENTVDHLLHAAHEHTDRLVLEFLA
GIESNCDAARKLGDIVCDLYALSVKARHGTSTIDTSPFSALKAIAGKSRMPMSAAL
CRDVRFPFSQSSCVTPRCRTGRICESELOPR"
complement(4959..5480)
/gene="ML0846"
/pseudo
complement(4959..5480)
/gene="ML0846"
/notes="Possible pseudogene similar to M. tuberculosis
paralogue Rv1747 (Best blastx score 127)"
/pseudo
/transl_table=11
/product="ABC transporter (pseudogene)"
complement(5522..5848)
/gene="ML0847"
/pseudo
complement(5522..5848)
/gene="ML0847"
/notes="Possible pseudogene of M. tuberculosis orthologue
Rv2327 (Best blastx score 184)"
/pseudo
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein (pseudogene)"
5905..8080
/gene="ML0848"
5905..8080
/gene="ML0848"
/notes="Similar to Mycobacterium tuberculosis hypothetical
ABC transporter ATP-binding protein Rv2326c or MTCY3G12.08
SW:YN26 MYCTU (P71886) (697 aa) fasta scores: E(): 0,
76.9% id in 697 aa. Shares similar domains with many
ABC-type transporters e.g. Streptomyces roseofulvus ATPase
component of putative ABC transporter frnD TR:068910
(EMBL:AF058302) (524 aa) fasta scores: E(): 8.4e-14, 31.5%
id in 495 aa and Synecococcus sp. nitrate transport
ATP-binding protein NrtD nrtD SW:NRTD SYN7 (P38046) (274
aa) fasta scores: E(): 1.6e-12, 34.8% id in 198 aa.
Previously sequenced as TR:O32971 (EMBL:Z98741). Contains
hydrophobic, possible membrane-spanning regions. Contains
2 Pfam matches to entry PF00017 ATP/GTP-binding site motif A
(P-loop). Contains 2 x PS00211 ABC transporters family
signature."
/codon_start=1
/transl_table=11
/product="ABC transporter"
/protein_id="CAC31229.1"
/db_xref="GI:13092926"
/db_xref="GOA:Q9CCF9"
/db_xref="SPTREMBL:Q9CCF9"
/translation="MDLCIHRMALLQCATRCHRWKLTQLCRAAPLADPGTGTIGKV
TADPRHLHRSGLQPELAQASVLAALCAVTAIVSVVVFPAAGLALGTVPMGLLA
YRYRFRULMNVAAAGVIAFLITGLGFIATVNSAYIGLTVGVKRGQGTLTVALA
LFAGLAAUNAVVALVIGLRHLIFKAMTANVDGIAILTMWHLFWAVOLKRYFAD
GLQHPWMLLGYFVITILVSLIGMWLVSLRVERIDIPDKVLDAFSAACNEDAPVGP
VPLWLDKVRFPYFHAGQDALREVSLDRVGEHVATVGTANGSKTTLMLILAGREPTSG
TVDRPGAVLQKLGKGTAVLQHPESQVLTGVADVVWGLPPGTDVDVNRLLREVRSGL
AFASRDTSLSGGELQRLAALAREPSLIIADEVTSMVDQCRDALLGLVSLGTRK
HPIALVHTHYNNEDATDTINLSDSPDNAGMAETVAPVSTVADVHRHPVLELV
CVGHEYSGTTPWAKAALHDSIFVRQSGDVLVYGSNGSGKSTLAWIAGLMVPTTGAC
LIDGRPTHEHVGAVALSFOAARLQLMRSRVLEVASRAGFSFRDREDRVAALGVVGLD

```

Query Match 74.6%; Score 19.4; DB 1; Length 348450;
 Best Local Similarity 95.2%; Pred. No. 4.4e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCCGCAATCTA 22
 |||||
 Db 95993 CAACGGCGAGCCCGCAACTA 95973

RESULT 10
 LOCUS AR226916 505 bp DNA linear PAT 20-DEC-2002
 DEFINITION Sequence 376 from patent US 6444799.
 ACCESSION AR226916
 VERSION AR226916.1 GI:27265489
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 505)
 AUTHORS Ross,B.C.
 TITLE P. Gingivalis polynucleotides and uses thereof
 JOURNAL Patent: US 6444799-A 376 03-SEP-2002;
 FEATURES Location/Qualifiers
 source 1..505
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Query Match 71.5%; Score 18.6; DB 6; Length 505;
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCCGCAATCTACGAA 26
 |||||
 Db 336 CAACGGCATGCGCGCAATCAACGAA 360

RESULT 11
 LOCUS Arabidopsis thaliana At3g02610 mRNA linear PLN 26-AUG-2003
 DEFINITION Arabidopsis thaliana complete cds.
 ACCESSION BT010447
 VERSION BT010447.1 GI:34222061
 KEYWORDS FLI CDNA
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 REFERENCE 1 (bases 1 to 1236)
 AUTHORS Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P., Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A., and Ecker,J.R.
 TITLE Arabidopsis ORF clones
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1236)
 AUTHORS Shinn,P., Chen,H., Cheuk,R., Kim,C.J., Bowser,L., Carninci,P., Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Shinozaki,K., Davis,R.W., Theologis,A., and Ecker,J.R.
 TITLE Direct Submission
 JOURNAL Submitted (26-AUG-2003) Salk Institute Genomic Analysis Laboratory (SIGAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

COMMENT
 RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-Length cDNA'). Seki M., Narusaka M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.
 The Salk, Stanford, PGEC (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAFL cDNAs: Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Bowser, L., Chan, M.M., Chang, C.M., Dale, J.M., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yamada, K., Yu, G., Davis, R.W., Theologis, A., and Ecker, J.R.
 Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.
 FEATURES source Location/Qualifiers
 1..1236
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="U61818"
 /ecotype="Columbia"
 /note="This clone is in pUNI 51"
 1..1236
 /note="acyl-[acyl-carrier-protein] desaturase (stearyl-ACP desaturase), putative"
 /codon_start=1
 /product="At3g02610"
 /protein_id="AA02867.1"
 /db_xref="GI:34222062"

CDS
 1..1236
 /translation="MALLNSTITIVAMKQNPVAVSPRTTCGSSPSPRLRLRVSCV
 ATNPSKTSEETKKKFPPIKEVNVTHITQEKLEIFKSMENACENLLSYLKPEVA
 SWQPDFELPNTDREDFEQVKELRDKETIPDIFYFVLVDMITBEALTYQTTLT
 LDGVKDTGSLTSMVAVRWAEENRHDGLNKLKLYLSGRVDMRHEKTIQVLIGF
 GMDSKFENPNYNGPIYTSFOERATFI SHGNTAKLATYGTTLAKICGTAADEKHE
 TATYRIVEKLEIPDPDGTGVALASMKRITMPAHLMDGDDDDLDPHYAAVQRIGV
 TATDYAGILEFLRLRWEKGLSGELSGEGRRAQDYLTLPQIRRLERANDRVKLA
 SKSPSVSFSNVIYGREVEL"

ORIGIN
 Query Match 71.5%; Score 18.6; DB 8; Length 1236;
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCCGCAATCTACGAA 26
 |||||
 Db 339 CAACGACGAGCAGCCGATCTACGAA 363

RESULT 12
 LOCUS Arabidopsis thaliana At3g02610 mRNA linear INV 19-DEC-2003
 DEFINITION Arabidopsis thaliana full insert cDNA.
 ACCESSION BT011075
 VERSION BT011075.1 GI:40215457
 KEYWORDS FLI CDNA
 SOURCE Arabidopsis thaliana (fruit fly)
 ORGANISM Arabidopsis thaliana
 REFERENCE 1 (bases 1 to 1766)
 AUTHORS Stapleton,M., Brokstein,P., Hong,L., Achavani,A., Carlson,J., Champs,R., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E., George,M., Gonzalez,M., Guarin,H., Krommiller,B., Li,P., Liao,G., Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M., and Celniker,S.
 TITLE Direct Submission
 JOURNAL Submitted (19-DEC-2003) Berkeley Drosophila Genome Project,

COMMENT
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.

FEATURES
source
1..1766
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
gene
1..51766
/gene="SRPK"
/db_xref="FLYBASE:FBgn0026370"
372..>1766
/notes="Longest ORF"
/codon_start=1
/product="SP09672p"
/protein_id="AAR82740.1"
/db_xref="GI:40215458"
CDS
/db_xref="FLYBASE:FBgn0026370"
/translation="MNKADYNERVLAIOAKKKHKNRKEKGKONGTNPQNSNQN
PFPSSNPFEPRNRSNENAPDANNHNNAESSAFKATATATASRSSGSGAGTPD
QVPPKTKAKPNTQKHQQPPRSSNYESSETPSDNEQDMEDICKGGHYVP
NIGLDPDRHYHVRKLGWGHFSTWLCWDLQAGVYVAIKIVSAPHAETADEKIL
KTVREDPNRRHKTVQMLDDPKITGVNGTHICMVFEVLGNLLKLRKSNRGIPL
ANVKTITROLEGLDYLVHCCIKIHTDIKPNVLYCVDPEPHVSRLATATQYCNWSK
MYPSLVSRAPKEVREBPITGKMSKNRKKLKKKAKRWELFKQRDYLEFQADGQGAN
PNEVONGDAGISDADEVFANSVEDNVHVAQPSKQREERKAPQESEEAQEDPL
TEGTDRAKKKKKKK"

ORIGIN
Query Match 71.5%; Score 18.6; DB 3; Length 1766;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGCGCCGAATCTACGA 25
Db 1561 AGAACGGCGAGCGCCGAATCTCCGA 1585

RESULT 13
AB046101/c
LOCUS
DEFINITION
Macaca fascicularis brain cDNA, clone:QcCE-16232.
ACCESSION
AB046101
VERSION
AB046101.1 GI:9280278
KEYWORDS
fis (full insert sequence).
SOURCE
Macaca fascicularis (crab-eating macaque)
ORGANISM
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutharia; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.
REFERENCE
1 (sites)
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hitai, M.,
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
Isolation of full-length cDNA clones from macaque brain cDNA

Libraries
Unpublished
2 (bases 1 to 1959)
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (14-JUL-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
URL: http://www.nih.go.jp/yoken/genebank/
Lib Name: macaque brain cDNA library QcCE
Lab host: TOR10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CACCATGTG)
Description: 1st strand cDNA was primed with an oligo(dt) primer
(ANGTGGCTTTTTTTTTTTTT); double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al. (University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
(5' end primer [CTTCGCTCTAAGAGCTGGC];
3' end primer [CGACCTGAGCTCGACACA]).
Location/Qualifiers
1..1959
/organism="Macaca fascicularis"
/mol_type="mRNA"
/db_xref="taxon:9541"
/clone="QcCE-16232"
/dev_stage="adult"
1034..1369
/codon_start=1
/product="unnamed protein product"
/protein_id="BAB01583.1"
/db_xref="GI:9280278"
/translation="MLVINTASVLLNFTILFYPENPLFLVLDGLSVVPELGLQLL
CSAEVFTSLCNHRSPSVLSLMTVLIGDKTRQLLGMNVGVAVGHVCVCTGICRPV
CKLATLLEV"

ORIGIN
Query Match 71.5%; Score 18.6; DB 9; Length 1959;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGGCGAGCGCCGAATCTACGA 25
Db 1140 ACAACTGAGAGGCCAGAACTAAGA 1116

RESULT 14
BD180518
LOCUS
DEFINITION
Highly thermophilic bacterium-derived protein and gene encoding it.
ACCESSION
BD180518
VERSION
BD180518.1 GI:30791436
KEYWORDS
JP 2002325574-A/1009.
Thermus thermophilus
SOURCE
Thermus thermophilus
ORGANISM
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE
1 (bases 1 to 2067)
Kiramitsu, N. and Yokoyama, S.
Autors
Highly thermophilic bacterium-derived protein and gene encoding it
Patent: JP 2002325574-A 1009 12-NOV-2002;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
JOURNAL
OS Thermus thermophilus
COMMENT
PN JP 2002325574-A/1009
PD 12-NOV-2002

PF 23-FEB-2001 JP 2001116171
 PI NARUKI KURAMITSU, SHIGEKI YOKOYAMA
 PC C12N15/09, C12N15/09, C07K14/195, C12N1/15, C12N1/19, C12N1/21, PC
 C12N5/10,
 PC C12N9/88, C12P21/02, (C12N9/88, C12R1:01), (C12N15/09, C12R1:01),
 PC (C12P21/02, C12R1:01), C12N15/00, C12N5/00, (C12N15/00,
 PC C12R1:01)
 CC Highly thermophilic bacterium-derived protein and gene CC
 encoding it
 FH Key Location/Qualifiers
 FT CDS (1)..(2067)
 FEATURES
 source
 1..2067
 Location/Qualifiers
 /organism="Thermus thermophilus"
 /mol_type="genomic DNA"
 /db_xref="taxon:274"

ORIGIN
 Query Match 71.5%; Score 18.6; DB 6; Length 2067;
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGAATCTACGA 25
 |||||
 DB 1472 AAAACGGCGAGCGCTGTTCTACGA 1496
 |||||

RESULT 15
 LOCUS AF301149 2368 bp mRNA linear INV 21-FEB-2001
 DEFINITION Drosophila melanogaster SR protein kinase 1 mRNA, complete cds.
 ACCESSION AF301149
 KEYWORDS
 VERSION AF301149.1 GI:10242346
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 2368)
 AUTHORS Allemand, E., Gattori, R., Bourbon, H.M., Stevenin, J., Caceres, J.F.,
 Soret, J. and Tazi, J.
 TITLE Distinctive features of Drosophila alternative splicing factor RS
 domain: implication for specific phosphorylation, shuttling, and
 splicing activation
 JOURNAL Mol. Cell. Biol. 21 (4), 1345-1359 (2001)
 MEDLINE 21106380
 PUBMED 11158320
 REFERENCE 2 (bases 1 to 2368)
 AUTHORS Allemand, E. and Tazi, J.
 TITLE Characterization of the Drosophila ortholog of the human SRPK1
 protein kinase
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 2368)
 AUTHORS Allemand, E. and Tazi, J.
 TITLE Direct Submission
 JOURNAL Submitted (29-AUG-2000) Centre Nationale de Recherche Scientifique,
 Institut de Genetique Moleculaire de Montpellier UMR-5535, IFR24,
 1919, Route de Mende, Montpellier, Herault 34293, France
 Location/Qualifiers
 1..2368
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /db_xref="taxon:7227"
 /chromosome="2"
 /map="52A1-52A2"
 1..2295
 /note="SRPK1; similar to Homo sapiens SRPK1"
 /codon_start=1
 /product="SR protein kinase 1"
 /protein_id="AAG15387.1"
 /db_xref="GI:10242347"
 /translation="MNKADVNRRLVLAIQAKKHKRKGKNGTNPQEQSNSSNQCN"

KPSSNPPEPNSRSENLOAPDNANNHLPSSAAEAKATATATATASTSGSGSAGTDP
 QIQPKIOAKAKENTQKQQQPPRSSNESYSEISFSDNEDELMDICKGGYHVP
 NIGLDFHRYHVIRKLGWGHSTVWLCWQAMGYVAIKVKSAPHFAETARDIKIL
 KIVRETFDSNPRRHKTQMLDDFKITGVNGVTHICWFEVLGDNLLKIRSNRYGIP
 ANVTITRVLEGLDYHTCKKIHTDKIPENVLLCVDEPHVRLATEATQLYCMNSK
 MYSLSVAPKEYREPTTGMKNKKLKKAKRMELFKQORDYLEADGOGAIN
 FNEVONGDAGISDADEVFANSVEDNVHVAQPSRKORDERKAPQPSREARQDPL
 TEGTDKAKKKKKKKKSKOSKQCPQLENSTSSAESKLSKQOANGSENSTNNK
 SNTSSGLKCKQSGKKKPLPKPKOKKQKHLNNNSKPNNSDSKISSGSEVENTSS
 ATNGHNSNTLPTPPPPKAKKAKKOPALDECNHVHVIADLGNACWDRHFTDIOT
 RQYNSLEVIIGAGYNTSADWSTACWPELATGYLPEPHSGESYTDREHLAIHEL
 LGPIPREILLNGTYAAKFTSCSLRNLISGLKPMGLMDVLEKYEWSQKDAASAPSEL
 TPMLEFDPNKRATAAECLQHPWLR"

ORIGIN
 Query Match 71.5%; Score 18.6; DB 3; Length 2368;
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGAATCTACGA 25
 |||||
 DB 1190 AGAACGGCGAGCGCGAATCTCCGA 1214
 |||||

RESULT 16
 LOCUS LMFLUNK03 40970 bp DNA linear INV 29-OCT-2000
 DEFINITION Leishmania major Friedlin cosmid L8530A, unmapped.
 ACCESSION AL449123
 VERSION AL449123.1 GI:11071675
 KEYWORDS
 SOURCE Leishmania major
 ORGANISM Leishmania major
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 REFERENCE 1 (bases 1 to 40970)
 AUTHORS Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and
 Smith, D.F.
 TITLE A physical map of the Leishmania major Friedlin genome
 JOURNAL Genome Res. 8 (2), 135-145 (1998)
 MEDLINE 98146435
 PUBMED 9477341
 REFERENCE 2 (bases 1 to 40970)
 AUTHORS Zimmermann, W., Wambutt, R., Ivens, A.C., Quail, M., Rajandream, M.A.
 and Barrell, B.G.
 TITLE Direct Submission
 JOURNAL Submitted (26-OCT-2000) European Leishmania major Friedlin genome
 sequencing project, Sanger Centre, The Wellcome Trust Genome
 Campus, Hinxton, Cambridge CB10 1SA, (E-mail: barrell@sanger.ac.uk)
 and Agowa GmbH, Glienickestrasse 185, D-12489, Berlin, Germany
 see <http://www.ebi.ac.uk/parasites/leish.html>
 Notes:
 Details of leishmania sequencing at the Sanger Centre are available
 on the World Wide Web.
 see <http://www.sanger.ac.uk/Projects/Lmajor/>
 CDS are numbered using the following system eg L8530.01. L8530
 (cosmid name), 01 (first CDS)
 To make the cosmid library Leishmania major Friedlin DNA was
 partially digested with Sau3AI prior to cloning into BamHI site of
 the cosmid shuttle vector CLHYG (Ryan et al. 1993 Gene
 131:145-150). The sequence of the packaged vector was determined by
 Peter Myler and Ken Stuart at Seattle Biomedical Research
 Institute, and is available as accession number U59231.
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous. The length
 in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for
 CDS which show significant similarity to other CDS in the database.
 Gene prediction is done using:
 (1)
 The Frameplot program of Bibb et al.,
 Gene 30:157-166(1984) as implemented
 at <http://www.nih.gov/jp/~jun/cgi-bin/frameplot.pl>. (2)

codon preference based on the codon usage table for Leishmania at <http://www.kazusa.or.jp/codon/>

(3) the Hexamer program which was written by Richard Durbin as an integral part of the ACEDB-based analysis tools for the C.elegans Genome Sequencing Project. The program calculates the log-likelihood score for a given DNA segment based on the frequency of 6-mers, normalised for the base-pair composition of the genome. The program was trained using a fasta file of confirmed Leishmania major coding sequences (CDS), i.e. from ATG start codon to the stop codon.

CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg) which is preceded by a stretch of pyrimidines or part of a kozak sequence. If this cannot be identified we choose the most upstream initiation codon. Transmembrane domains were predicted as implemented at the TMHMM server: <http://www.cbs.dtu.dk/services/TMHMM-1.0/>. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid L8530A was thought to map to chromosome 5, but does not, and is currently not overlapped by any cosmid sequenced to date.

FEATURES

```

source
1..40970
    /organism="Leishmania major"
    /mol_type="genomic DNA"
    /strain="Friedlin"
    /db_xref="taxon:5664"
    /clone="cosmid L8530A"
1..174
    /gene="L8530.01"
<1..174
    /gene="L8530.01"
    /note="L8530.01, len > 56 aa, unknown; good similarity to Q9SVU9, hypothetical 22.5 kd protein (206 aa, Arabidopsis thaliana, EMBL: AL161573, CAB81470); Pasta scores: E(1) 0.07, 42.0% identity in 50 aa"
    /codon_start=1
    /label="L8530.01"
    /product="conserved hypothetical protein"
    /protein_id="CAC14604.1"
    /db_xref="GI:11071676"
    /db_xref="SPTREMBL:O9GRN5"
    /translation="DPRVRSALRSPSLQKLIKIDSRSLDALEAAQYNNADFKGFC DAVIRVIAEVEGR"
    /complement(269..283)
    /note="poly-pyrimidine tract"
789..804
    /note="poly-pyrimidine tract"
792..803
    /note="(cct)4"
908..922
    /note="poly-pyrimidine tract"
1185..1202
    /note="poly-pyrimidine tract"
1273..1294
    /note="poly-pyrimidine tract"
    /complement(1490..1506)
    /note="poly-pyrimidine tract"
1507..1522
    /note="(tgg)4"
1727..1814
    /note="region of BLASTN similarity to: 15..102 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 77% identity over 87 bases"
    /note="region of BLASTN similarity to: 18934..18923 AL132764 Leishmania major Friedlin chromosome 4 PAC P1421, 72% identity over 89 bases"
    /note="region of BLASTN similarity to: 29815..29904 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 72% identity over 89 bases"
    /note="region of BLASTN similarity to: 45..112 AL139794"

```

```

Leishmania major Friedlin chromosome 4 PAC P1105, 86% identity over 67 bases
    /note="region of BLASTN similarity to: 27910..27986 AL359781 Leishmania major Friedlin chromosome 21 cosmid L3302, 79% identity over 76 bases"
    /note="region of BLASTN similarity to: 7374..7431 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 84% identity over 57 bases"
    /note="region of BLASTN similarity to: 24835..24899 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 81% identity over 64 bases"
    /complement(1733..1801)
    /note="region of BLASTN similarity to: 5998..6066 AL034357 Leishmania major Friedlin chromosome 4 cosmid L2969, 76% identity over 68 bases"
    /note="region of BLASTN similarity to: 2900..2990 AC004145 Leishmania major chromosome 3 clone L5801 strain Friedlin, complete sequence, 72% identity over 90 bases"
    /note="region of BLASTN similarity to: 34910..34987 AL139794 Leishmania major Friedlin chromosome 4 PAC P1105, 86% identity over 57 bases"
    /note="region of BLASTN similarity to: 3605..3700 AC004145 Leishmania major chromosome 3 clone L5801 strain Friedlin, complete sequence, 81% identity over 95 bases"
    /note="region of BLASTN similarity to: 284..378 AL161414 Leishmania major Friedlin chromosome 19 cosmid L5515, 80% identity over 94 bases"
    /note="region of BLASTN similarity to: 282..370 AL161414 Leishmania major Friedlin chromosome 19 cosmid L5515, 82% identity over 88 bases"
    /note="region of BLASTN similarity to: 25755..25841 AL034357 Leishmania major Friedlin chromosome 4 cosmid L2969, 82% identity over 86 bases"
    /note="region of BLASTN similarity to: 33949..34050 AL359773 Leishmania major Friedlin chromosome 14 PAC P1295, 73% identity over 101 bases"
    /note="region of BLASTN similarity to: 25749..25840 AL034357 Leishmania major Friedlin chromosome 4 cosmid L2969, 82% identity over 91 bases"
    /note="region of BLASTN similarity to: 3379..3460 AC004145 Leishmania major chromosome 3 clone L5801 strain Friedlin, complete sequence, 76% identity over 81 bases"
    /note="region of BLASTN similarity to: 33726..33826 AL359773 Leishmania major Friedlin chromosome 14 PAC P1295, 68% identity over 100 bases"
    /note="region of BLASTN similarity to: 25744..25841 AL034357 Leishmania major Friedlin chromosome 4 cosmid L2969, 80% identity over 97 bases"
    /note="region of BLASTN similarity to: 3591..3701 AC004145 Leishmania major chromosome 3 clone L5801 strain Friedlin, complete sequence, 76% identity over 110 bases"
    /note="region of BLASTN similarity to: 24557..24649 AC009605 Leishmania major chromosome 35 clone L8019 strain Friedlin, complete sequence, 73% identity over 92 bases"
    /note="region of BLASTN similarity to: 236017..236106 AF008205 Leishmania major chromosome 1, complete sequence, 75% identity over 89 bases"
    /note="region of BLASTN similarity to: 1548..1649 AL049188 Leishmania major Friedlin chromosome 4 cosmid L5683, 66% identity over 101 bases"
    /complement(1860..2342)
    /note="region of BLASTN similarity to: 1..483 AQ850336 LMAJFV1 lm35e04.x1 Leishmania major FV1 random genomic library Leishmania major genomic clone LMAJFV1 lm35e04.3, similar to contains element 284-bp.4 Leishmania repetitive element; 99% identity over 482 bases"
    /note="poly-pyrimidine tract"
1910..1925
    /complement(2583..2655)
    /note="region of BLASTN similarity to: 32884..32956 AF239995 Leishmania major strain Friedlin chromosome 5 cosmid L8530, complete sequence, 78% identity over 72 bases"

```

misc_feature

misc_feature

repeat_region

misc_feature

region of BLASTN similarity to: 3285. .33003 AF239995
 Leishmania major strain Friedlin chromosome 5 cosmid
 L8530, complete sequence, 68% identity over 118 bases
 region of BLASTN similarity to: 9686. .9748 AF239995
 Leishmania major strain Friedlin chromosome 5 cosmid
 L8530, complete sequence, 82% identity over 62 bases
 region of BLASTN similarity to: 22305. .22440 AC009605

Query Match 71.5%; Score 18.6; DB 3; Length 40970;
 Best Local Similarity 84.0%; Pred. No. 1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCGAATCTACGAA 26
 |||||
 DB 1349 CACGGCGAGCGCGAATCAACGAA 1325

RESULT 17
 AC017471 42950 bp DNA linear HTG 09-DEC-1999
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 ACCESSION AC017471
 VERSION AC017471.1 GI:6553515
 KEYWORDS HTG; HTGS PHASE2.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 42950)
 AUTHORS Adams.M. and Venter.J.C.
 TITLE Direct Submission

JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

COMMENT This sequence was identified as CDM:10210907 by the submitter.
 For more information on this record e-mail to fly@cblera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..42950
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"

ORIGIN

Query Match 71.5%; Score 18.6; DB 2; Length 42950;
 Best Local Similarity 84.0%; Pred. No. 1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCGAATCTACGA 25
 |||||
 DB 10972 AGAACGGCGAGCGCGAATCTCCA 10996

RESULT 18
 ATAC021640 103904 bp DNA linear PLN 30-OCT-2002
 LOCUS Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence,
 DEFINITION complete sequence.

ACCESSION AC021640
 VERSION AC021640.7 GI:12408747
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 103904)

AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.,
 Wu,D., Rensing,C.M., Koo,H., Fujii,C.Y., Otterback,T.R., and
 Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence
 2 (bases 1 to 103904)
 Lin,X. and Kaul,S.
 Direct Submission
 Submitted (18-JAN-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Lin,X.
 Direct Submission
 Submitted (24-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Jan 24, 2001 this sequence version replaced gi:12280770.
 Address all correspondence to:at@tigr.org
 BAC clone F16B3 is from Arabidopsis chromosome III and is near the
 molecular marker mi74.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL (available by anonymous ftp
 from arthur.sgm.cornl.gov); Genefinder (Phil Green, University of
 Washington); Genscan (Chris Burge, GENSCANW.html), and NetPlantGene
 (http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted as tRNAscan-SE (Sean Eddy,
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
 identified by RepeatMasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES
 source

1..103904
 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="mi74"
 /clone="F16B3"
 <129..>979
 /gene="F16B3.1"
 /note="Predicted by genefinder"
 join(<129..241,586..>979)
 /gene="F16B3.1"
 join(129..241,586..979)
 /gene="F16B3.1"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="AAF32445.1"
 /db_xref="GI:5957701"
 /translation="WEERWEGSPLENERAGLLCYDDEEYQFGPAPKLPFRYLSEIG
 LQLLGGDDVVISLKOLYSGIAEGYKGCWENYEVYKLGKGLYLGRHGVPMWTKY
 AVNTTPSDESLCAAEFFQDRDSVTKLLSDMHICDARPVFDVYLPNSQFKKSGPSP
 SFVTCFSG"

complement(<3079..>4265)
 /gene="F16B3.2"
 /note="identical to COL2 GB:AB67879 from [Arabidopsis
 thaliana]"
 complement(join(<3079..3438,3582..>4265))
 /gene="F16B3.2"
 complement(join(<3079..3438,3582..>4265))
 /gene="F16B3.2"
 /codon_start=1
 /product="COL2"
 /protein_id="AAF32446.1"

```
/db_xref="GI:6957702"
/translation="MLKEESNEGTWARACDTCRRAACTVYCEADSAVLCTTCDAVRH
AANVASRERVQCSCEBAPALCADAASLCTACDAEIHSAANLARRHORVIL
PLANSQCSARSETADNEDREVALNPNKNTGNONNGFLVGVYLDLVDYS
SMDNQFNOYTHVORSFGGVVLPLOVEESTSLQSQOQFOLGNGPSSGAHYN
NSLKLHNSAVSSMDISVPESATSDITVQHPPTTKETIDQSGPQVQOOLTPM
REARVLRYREKKRKDKPTIYASRAKAYAIRPKRFAKRIETAEAEIEFSTS
LMSGTGIVGVPSP"
repeat_region
5228..5255
/db_pt_family="(CAA)n"
<5471..>7228
/genes="F16B3.3"
/notes="predicted by genefinder"
<5471..>7228
/genes="F16B3.3"
5471..7228
/genes="F16B3.3"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF32447.1"
/db_xref="GI:6957703"
/translation="MVFPILKLDFTQGRAGDSLGKPGKSTIRIGRIVRGNIEIAKDA
GTSKHLRVSESENIWIHDLGSNGTILNSDTIDTPVNSHGEIKLGYSTILV
NFGSDVQAPQHLPPRRNNKRLAASDPDPIESVQEKPKTRGSSKQSENEELP
KSTRASKNLDDIADKEBELDVEIEKVVKARVGRPKNGSAIAKEBESVKEKRVG
RKNKASALTEBEEVVEKGNRRARGNSEIYQKSLKLEVEDTPKAVEI SEVKS
KRVTRSKQENECFGEVDEKATKTRSKTIEIGESFLEIEMVLNQARKSRAKX
KMDPESTRNDAGEEVLKNCHEEDKNEAQCGSGRSDDICQDEKSCDGSKR
VEQVEIELKSTVEGLKCTDKEDGETENSQDIERENENKEAQCGCSRDEKYE
RVGGAKRVQVEI ELRKSTVGEDDLNCTVREDGTENLQIEEBCHDESDCKVEE
AGPATUDEKVGQGNKVKGRVEVYELKMLREWPDFAEVQLPKQITBETKMIETPM
RKSMSVHKHIAQEKQDP"
complement(<7759..>10518)
/genes="F16B3.4"
/notes="predicted by genscan, contains Coesterase domain"
complement(join(<7759..7920,8001..8091,8192..8335,
8796..8861,8964..9101,9182..9283,9433..9507,9638..9718,
10256..>10518))
/genes="F16B3.4"
complement(join(7759..7920,8001..8091,8192..8335,
8796..8861,8964..9101,9182..9283,9433..9507,9638..9718,
10256..>10518))
/genes="F16B3.4"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF32448.1"
/db_xref="GI:6957704"
/translation="MQLSPRCRPMSENRANSSEEMELLHGNSRLSSPHEVRRV
SGNSSEGSRIQCSFGDRIGHAAETIILRLSLGIDLYIPPTSDGLKP
VYVFTVGAWIIGYKAWGSLIGLAERDIIVACLDYRNFPOQTSDMSVDAQHSIF
VCNNIGAFGDPNRIYLMGOSAGAHISSCALFEQAIKESRGSISMSVSQIKAYFGLS
GGYNLNLVPHFNRLGYSIFLSIMEGESPKQSPFVRLKDLNVRKAAALLPHIIL
FHGSADYSIPPEASKYFTDALQAAEVKAEIYVYKGTHTDLFLQDPLRGKGLDFDHI
VSMIHADSDALRNDVAPRKLRLVPEFLIKLAGRVSPF"
complement(11800..14456)
/genes="F16B3.5"
complement(join(11800..11972,12355..12394,12527..12640,
12750..12834,12919..13039,13133..13258,13559..13659,
13745..13874,14237..14456))
/genes="F16B3.5"
complement(join(11820..11972,12355..12394,12527..12640,
12750..12834,12919..13039,13133..13258,13559..13659,
13745..13874,14237..14413))
/genes="F16B3.5"
/notes="unknown protein"
/codon_start=1
/protein_id="AAF32449.1"
/db_xref="GI:6957705"
/translation="MAEGEDSQRLLKIAAAYDENDARWADYWNILIPPHASRP
EVDVHFRKRYQYIDPLVPEPMSSTSSSSQARPTATASASANNQVRNS
GYSVPTSGSATGAPSSMRWDEQIQFVNAWFIYAVLPIPKNLNRLK
SFWGTACSSLYSLYSGRPRAWNGLOVYFQSIYAAKDFIYICLTFVTSHLCLK
FALIPILCRALEQAVKELRNFGRSTIYRYLEDPCVWVESNTTILNLSQAEIAG
```

```
FLIISLLSNQRNIIQTFNFWQOLLKMLNYQAPVTAGYHQSTWSRIGRTVTPIQRYAPF
LNPVSVAQWRWER"
/genes="F16B3.6"
/notes="predicted by genefinder"
<15001..>15660
/genes="F16B3.6"
15001..15660
/genes="F16B3.6"
15001..15660
/genes="F16B3.6"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF32450.1"
/db_xref="GI:6957706"
/translation="MSALRLRNANTPAPELDELSDOTPSRVLKROMSQRAMSN
LTSANLSNLLPTGTLIAQLQLLPFTFNSGVCDHATRELTAVLLFLLAASCFVSSFTD
SVKADDTGTYFGVTFKGMWVDYDFSGGLGDLPAKIRMFVDMWIHATLSLVFGAV
R"
complement(<16445..>17870)
/genes="F16B3.7"
/notes="predicted by genscan"
complement(join(<16445..16464,16657..17107,17220..>17870))
/genes="F16B3.7"
complement(join(16445..16464,16657..17107,17220..17870))
/genes="F16B3.7"
/notes="hypothetical protein"
/codon_start=1
/protein_id="AAF32451.1"
/db_xref="GI:6957707"
/translation="MELPATKRLRGLVIFPILITIAPILYFGYELYSTYKHL
NSSSSSAISSPSSYHNSSSSDSYKTEDSPSYDNDYDTHDPSSSLUHNDRLS
ISSNGHQVTPKKEHRRKKRKCIDIFSGEWINPKAPYNTTCRAIHEHQNCIKY
GRPLGFMKWKWKCDLPLFDPEYLEIIVRGTRMAFVGDSVSRNHVQSILCLLSRV
EHPGDSQOEENFORWKYNTYETIATFTWTHLVRABETGTGPTGPNSSFYNYLDEPD
PTWASQGEFEYIILISSQWFEFPLEFDKCRIGCLCYIIPGVNVAHFAYRRALR
TTFKILGLENFKEVFLRTFAPSHFEGGEGWDKGMQRQS"
complement(11673..17778)
/rpt_family="(CAT)n"
complement(18048..18074)
Query Match 71.5%; Score 18.6; DB 8; Length 103904;
Best Local Similarity 84.0%; Pred. No. ie+03; 4; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;
QY 2 CAACGGCGAGGCCGCAATCTACGAA 26
|||||
Db 72110 CAACGACGACGACGATCTACGAA 72134
|||||
RESULT 19
LMFLCHR31_13
WPCOMMENT
Sequence split into 22 fragments LOCUS LMFLCHR31 Accession AL499621
Fragment Name Begin End
LMFLCHR31_00 1 110000
LMFLCHR31_01 100001 210000
LMFLCHR31_02 200001 310000
LMFLCHR31_03 300001 410000
LMFLCHR31_04 400001 510000
LMFLCHR31_05 500001 610000
LMFLCHR31_06 600001 710000
LMFLCHR31_07 700001 810000
LMFLCHR31_08 800001 910000
LMFLCHR31_09 900001 1010000
LMFLCHR31_10 1000001 1110000
LMFLCHR31_11 1100001 1210000
LMFLCHR31_12 1200001 1310000
LMFLCHR31_13 1300001 1410000
LMFLCHR31_14 1400001 1510000
LMFLCHR31_15 1500001 1610000
LMFLCHR31_16 1600001 1710000
LMFLCHR31_17 1700001 1810000
LMFLCHR31_18 1800001 1910000
```

LMFLCHR31_19 1500001 2010000
 LMFLCHR31_20 2000001 2110000
 LMFLCHR31_21 2100001 2117963
 Continuation (14 of 22) of LMFLCHR31 from base 1300001 (AL499621 Leishmania major chromosome 2)

Query Match 71.5%; Score 18.6; DB 2; Length 110000;
 Best Local Similarity 84.0%; Pred. No. 1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26

DB 24990 CACGGCGAGCGCGCAATCTACGAA 25014

RESULT 20

LMFLCHR32_17/c

WPCOMMENT

Sequence split into 28 fragments LOCUS LMFLCHR32 Accession AL499622

Fragment Name	Begin	End
LMFLCHR32_00	1	110000
LMFLCHR32_01	100001	210000
LMFLCHR32_02	200001	310000
LMFLCHR32_03	300001	410000
LMFLCHR32_04	400001	510000
LMFLCHR32_05	500001	610000
LMFLCHR32_06	600001	710000
LMFLCHR32_07	700001	810000
LMFLCHR32_08	800001	910000
LMFLCHR32_09	900001	1010000
LMFLCHR32_10	1000001	1110000
LMFLCHR32_11	1100001	1210000
LMFLCHR32_12	1200001	1310000
LMFLCHR32_13	1300001	1410000
LMFLCHR32_14	1400001	1510000
LMFLCHR32_15	1500001	1610000
LMFLCHR32_16	1600001	1710000
LMFLCHR32_17	1700001	1810000
LMFLCHR32_18	1800001	1910000
LMFLCHR32_19	1900001	2010000
LMFLCHR32_20	2000001	2110000
LMFLCHR32_21	2100001	2210000
LMFLCHR32_22	2200001	2310000
LMFLCHR32_23	2300001	2410000
LMFLCHR32_24	2400001	2510000
LMFLCHR32_25	2500001	2610000
LMFLCHR32_26	2600001	2710000
LMFLCHR32_27	2700001	2727709

Continuation (18 of 28) of LMFLCHR32 from base 1700001 (AL499622 Leishmania major chromosome 2)

Query Match 71.5%; Score 18.6; DB 2; Length 110000;
 Best Local Similarity 84.0%; Pred. No. 1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26

DB 80631 CACGGCGAGCGCGCAATCTACGAA 80607

RESULT 21

AC008343

LOCUS

DEFINITION Drosophila melanogaster, chromosome 2R, region 51F-52A, BAC clone INV 29-MAR-2001

ACCESSION AC008343

VERSION AC008343.4

KEYWORDS HTG.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 158402)

AUTHORS

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorset, V., Dou, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Housh, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, M., Jalali, M., Kruse, D., Li, P., Matti, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunco, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, P., Stapleton, M., Strong, R., Svirskaas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 158402)
 Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenoff, C., Champs, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R.A., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirskaas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

TITLE

JOURNAL

COMMENT

Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Mar 29, 2001 this sequence version replaced gi:6563415.
 Sequence submitted by:
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES

source

1. 158402
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="Y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /map="51F-52A"
 /clone="BACR03D24 (D931)"
 /clone_lib="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACE3.6)"

ORIGIN

Query Match 71.5%; Score 18.6; DB 3; Length 158402;
 Best Local Similarity 84.0%; Pred. No. 1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGAATCTACGAA 25

DB 34691 AGAACGGCGAGCGCCGAATCTCCGA 34715

RESULT 22

AC011127/c

LOCUS

DEFINITION Homo sapiens chromosome 18 clone RP11-110J22 map 18, WORKING DRAFT

SEQUENCE, 11 unordered pieces.

ACCESSION AC011127

VERSION AC011127.6

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K., McSheeters, R., Meldrum, J., Meneus, D., Morrow, J., Naylot, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE

Direct Submission
Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 169553)

REFERENCE

AUTHORS

Direct Submission
 Submitted (25-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (Bases 1 to 169553)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Brown,A., Canarata,J., Campopiano,A., Chang,J., Chazaro,B., Chospel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Lander,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meidrim,J., Meness,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (Bases 1 to 169553)
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Canarata,J., Chang,J., Chazaro,B., Chospel,Y., Collymore,A., Canarata,J., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lander,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meidrim,J., Meness,L., Mihova,T., Mlenga,V., O'Connor,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schnupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 17, 2002 this sequence version replaced gl:21307364.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L3356
 Center clone name: 13 L 22

TITLE

Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

REFERENCE

Ouyang, S., Zaidman, B., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 4 (bases 1 to 169553)
 Bitren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barina, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearillano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karatas, A., Kellis, C., Lander, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J. J., Matthews, C.,
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mleung, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J. J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Riser, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schubback, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, K., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zaidman, J.,
 Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 17, 2002 this sequence version replaced gi:21307364.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WBIR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L3156
 Center clone name: 13 L 22

TITLE

Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

WAVES

On Jun 17, 2002 this sequence version replaced gi:21307364.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RX/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBS
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3356
Center clone name: 13 L 22

```

repeat_region complement(14021..14065)
/rpt_family="L2"
repeat_region complement(14070..14121)
/rpt_family="(CGAGAA)n"
repeat_region complement(14204..14259)
/rpt_family="MIR"
repeat_region complement(14599..14911)
/rpt_family="AluSq"
repeat_region complement(16231..16231)
/rpt_family="(GGA)n"
repeat_region complement(16243..16315)
/rpt_family="AluJ/FLAM"
repeat_region complement(16561..17000)
/rpt_family="MLT1C"
repeat_region complement(17059..17174)
/rpt_family="MLT1C"
repeat_region complement(17180..17236)
/rpt_family="MLT1C"
repeat_region complement(18051..18366)
/rpt_family="AluSp"
repeat_region complement(18447..18749)
/rpt_family="AluSq"
repeat_region complement(19254..19367)
/rpt_family="MIR"
repeat_region complement(22413..22413)
/rpt_family="AT rich"
repeat_region complement(22681..22989)
/rpt_family="LINE4A"
repeat_region complement(23295..23295)
/rpt_family="AluX"
repeat_region complement(23296..23329)
/rpt_family="LINE4A"

```

Query Match 71.5%; Score 18.6; DB 9; Length 169553;
 Best Local Similarity 84.0%; Pred. No. 1e+03;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGCGAGGCGGAGTCTACA 25
 DB 163397 ACAACTGAGAGCGGAGTCTAAGA 163373

RESULT 24
 AC023532/c
 LOCUS 189412 bp DNA linear HTG 10-SEP-2000
 DEFINITION Homo sapiens chromosome 11 clone RP11-150F7 map 11, WORKING DRAFT
 SEQUENCE, 39 unordered pieces.
 ACCESSION AC023532
 VERSION AC023532.5 GI:10047802
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 189412)
 Barren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavskiy,L.,
 Bouckgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
 Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
 Fenesor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Landers,T., Larcocque,K., Lehoczy,J., Levine,R.,
 Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
 McEwan,P., McCurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
 Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,F.M.,

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
 Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Subramanian,A., Tallamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
 Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
 Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
 Zody,M.

TITLE
 JOURNAL
 COMMENT

Direct Submission
 Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 10, 2000 this sequence version replaced gi:7657721.
 All repeats were identified using RepeatMasker:

Smt, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5060

Center clone name: 150.F.7

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 15822 bases at least Q40

Consensus quality: 168315 bases at least Q30

Consensus quality: 177267 bases at least Q20

Insert size: 178000; agarose-fp

Insert size: 185612; sum-of-contigs

Quality coverage: 4.4 in Q20 bases; agarose-fp

Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 39 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 4164: contig of 4164 bp in length

* 4165: gap of 100 bp

* 4265: contig of 1071 bp in length

* 5336: gap of 100 bp

* 5436: contig of 1130 bp in length

* 6665: gap of 100 bp

* 7843: contig of 1178 bp in length

* 7944: gap of 100 bp

* 9118: contig of 1175 bp in length

* 9219: gap of 100 bp

* 9219: contig of 1157 bp in length

* 10376: gap of 100 bp

* 10476: contig of 1176 bp in length

* 11652: gap of 100 bp

* 11752: contig of 1305 bp in length

* 13057: gap of 100 bp

* 13157: contig of 1183 bp in length

* 13440: gap of 100 bp

* 14440: contig of 1294 bp in length

* 15734: gap of 100 bp

* 15834: contig of 2056 bp in length

* 17890: gap of 100 bp

* 17990: contig of 1056 bp in length

* 19046: gap of 100 bp

* 19145: contig of 1435 bp in length

* 20580: gap of 100 bp

* 20680: contig of 1812 bp in length

* 22492: gap of 100 bp

* 22592: contig of 1421 bp in length

* 24013: gap of 100 bp

* 24113: contig of 1688 bp in length

* 25801: gap of 100 bp

* 25801: contig of 1688 bp in length

* 26114: gap of 100 bp

* 26114: contig of 1688 bp in length

* 26114: gap of 100 bp

* 26114: contig of 1688 bp in length

* 26114: gap of 100 bp

* 26114: contig of 1688 bp in length

* 26114: gap of 100 bp

* 26114: contig of 1688 bp in length

* 26114: gap of 100 bp

* 26114: contig of 1688 bp in length

* 26114: gap of 100 bp

* 26114: contig of 1688 bp in length

* 26114: gap of 100 bp

* 26114: contig of 1688 bp in length

Assembly program: Phrap; version 0.990329
Consensus quality: 190215 bases at least Q40
Consensus quality: 193565 bases at least Q30
Consensus quality: 194985 bases at least Q20
Insert size: 195937; sum-of-contigs
Quality coverage: 9.11x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 22006 contig of 22006 bp in length
22107 42860 contig of 20754 bp in length
42961 63463 contig of 20503 bp in length
63564 81431 contig of 17868 bp in length
81532 99656 contig of 18125 bp in length
99757 116826 contig of 17070 bp in length
116927 132183 contig of 15257 bp in length
132284 142638 contig of 10355 bp in length
142739 152508 contig of 9770 bp in length
152609 161840 contig of 9232 bp in length
161941 169129 contig of 7189 bp in length
169230 175273 contig of 6044 bp in length
175374 180468 contig of 5095 bp in length
180569 183944 contig of 3376 bp in length
184045 187766 contig of 3722 bp in length
187867 190600 contig of 2734 bp in length
190701 193108 contig of 2307 bp in length
193108 194983 contig of 1876 bp in length
194983 196553 contig of 1369 bp in length
196553 197837 contig of 1285 bp in length.
NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 22006: contig of 22006 bp in length
22007 22106: gap of 100 bp
22107 42860: contig of 20754 bp in length
42861 42960: gap of 100 bp
42961 63463: contig of 20503 bp in length
63464 63564: gap of 100 bp
63564 81431: contig of 17868 bp in length
81432 81532: gap of 100 bp
81532 99656: contig of 18125 bp in length
99657 99756: gap of 100 bp
99757 116826: contig of 17070 bp in length
116827 116926: gap of 100 bp
116927 132183: contig of 15257 bp in length
132184 132284: gap of 100 bp
132284 142638: contig of 10355 bp in length
142639 142738: gap of 100 bp
142739 152508: contig of 9770 bp in length
152509 152608: gap of 100 bp
152609 161840: contig of 9232 bp in length
161841 161941: gap of 100 bp
161941 169129: contig of 7189 bp in length
169130 169230: gap of 100 bp
169230 175273: contig of 6044 bp in length
175274 175373: gap of 100 bp
175374 180468: contig of 5095 bp in length
180469 180568: gap of 100 bp
180569 183944: contig of 3376 bp in length
183945 184044: gap of 100 bp
184045 187766: contig of 3722 bp in length
187767 187866: gap of 100 bp
187867 190600: contig of 2734 bp in length

* 190601 190700: gap of 100 bp
* 190701 193007: contig of 2307 bp in length
* 193008 193107: gap of 100 bp
* 193108 194983: contig of 1876 bp in length
* 194984 195083: gap of 100 bp
* 195084 196452: contig of 1369 bp in length
* 196453 196552: gap of 100 bp
* 196553 197837: contig of 1285 bp in length.

FEATURES

source
1..197837
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q22"
/clone="RP11-825H23"
misc_feature
1..22006
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature
22107..42860
/note="assembly_fragment"
misc_feature
42961..63463
/note="assembly_fragment"
misc_feature
63564..81431
/note="assembly_fragment"
misc_feature
81532..99656
/note="assembly_fragment"
misc_feature
99757..116826
/note="assembly_fragment"
misc_feature
116927..132183
/note="assembly_fragment"
misc_feature
132284..142638
/note="assembly_fragment"
misc_feature
142739..152508
/note="assembly_fragment"
misc_feature
152609..161840
/note="assembly_fragment"
misc_feature
161941..169129
/note="assembly_fragment"
misc_feature
169230..175273
/note="assembly_fragment"
misc_feature
175374..180468
/note="assembly_fragment"
misc_feature
180569..183944
/note="assembly_fragment clone_end:T7 vector_side:left"
misc_feature
184045..187766
/note="assembly_fragment"
misc_feature
187867..190600
/note="assembly_fragment"
misc_feature
190701..193007
/note="assembly_fragment"
misc_feature
193108..194983
/note="assembly_fragment"
misc_feature
195084..196452
/note="assembly_fragment"
misc_feature
196553..197837
/note="assembly_fragment"

ORIGIN

Query Match 71.5%; Score 18.6; DB 2; Length 197837;
Best Local Similarity 84.0%; Pred No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGAATCTACGA 25
|||||
DB 125189 ACAACTGAGAGCCGAGATCTAAGA 125213

RESULT 26

AP002985/c 205152 bp DNA linear PRI 15-MAR-2003
LOCUS Homo sapiens genomic DNA, chromosome 11q clone:CTD-2542M24,
DEFINITION complete sequence.
ACCESSION AP002985

```

VERSION      AP002985.2  GI:16751492
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Homo sapiens genomic DNA
JOURNAL      Published Only in Database (2000)
AUTHORS      2 (bases 1 to 205152)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suhiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              (E-mail:hattori@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      On Nov 5, 2001 this sequence version replaced gi:11559300.
FEATURES     Location/Qualifiers
             source
               1..205152
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               /chromosome="11"
               /map="11q"
               /clone="CTD-2542M24"
ORIGIN
Query Match      71.5%; Score 18.6; DB 9; Length 205152;
Best Local Similarity 84.0%; Pred.NO.le+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CAACGGCGAGGCCCCGAATCTACGAA 26
        |||||
Db      50297 CAACAGCGAGGCCCCGTCTCTACAA 50273
        |||||

RESULT 27
AP001447
LOCUS      214701 bp DNA linear HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-705P3 map 11q14, WORKING
DRAFT SEQUENCE, 46 unordered pieces.
ACCESSION AP001447.2 GI:8117321
VERSION     HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 214701)
AUTHORS     Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Homo sapiens 214,701 genomic DNA of 11q14
JOURNAL     Published Only in Database (2000)
AUTHORS     2 (bases 1 to 214701)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (15-MAR-2000) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
              Japan (E-mail:hattori@sc.riken.go.jp,
              URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
              Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7262580.
             ----- Genome Center
             Center: RIKEN Genomic Sciences Center (GSC)
             Center code: RIKEN
             Web site: http://hgp.gsc.riken.go.jp/
             Contact: hattori@sc.riken.go.jp

```

```

----- Project Information
Center project name: HumDraft11
Center clone name: Rp11-705P3
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 182011 bases at least Q40
Consensus quality: 197127 bases at least Q30
Consensus quality: 205118 bases at least Q20
Insert size: 210201; sum-of-contigs
Quality coverage: 4.34x in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 46 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 14737 contig of 14737 bp in length
14838 28083 contig of 13246 bp in length
28184 40511 contig of 12328 bp in length
40612 51898 contig of 11287 bp in length
51999 64081 contig of 12083 bp in length
64182 74827 contig of 10646 bp in length
74928 84831 contig of 9904 bp in length
84932 91901 contig of 6969 bp in length
92001 99014 contig of 7014 bp in length
99115 105728 contig of 6614 bp in length
105829 110842 contig of 5014 bp in length
110943 116467 contig of 5525 bp in length
116568 121590 contig of 5023 bp in length
121691 127550 contig of 5860 bp in length
127651 132127 contig of 4477 bp in length
132228 136727 contig of 4500 bp in length
136828 141378 contig of 4551 bp in length
141479 145693 contig of 4215 bp in length
145794 149654 contig of 3861 bp in length
149755 154794 contig of 5040 bp in length
154895 158481 contig of 3587 bp in length
158582 164776 contig of 6195 bp in length
164877 167761 contig of 2885 bp in length
167862 171113 contig of 3252 bp in length
171214 174080 contig of 2876 bp in length
174181 177056 contig of 2876 bp in length
177157 179451 contig of 2295 bp in length
179552 182040 contig of 2489 bp in length
182141 184418 contig of 2278 bp in length
184519 186523 contig of 2005 bp in length
186624 188462 contig of 1839 bp in length
188563 190196 contig of 1634 bp in length
190297 192401 contig of 2105 bp in length
192502 194597 contig of 2096 bp in length
194698 198254 contig of 1557 bp in length
198355 198593 contig of 2239 bp in length
198694 200646 contig of 1953 bp in length
200747 202409 contig of 1663 bp in length
202510 204484 contig of 1975 bp in length
204585 206037 contig of 1453 bp in length
206138 207948 contig of 1811 bp in length
208049 209677 contig of 1629 bp in length
209778 210908 contig of 1131 bp in length
211009 212394 contig of 1386 bp in length
212495 213510 contig of 1016 bp in length
213611 214701 contig of 1091 bp in length

```

Sequence updated (26-May-2000).
 * NOTE: This is a 'working draft' sequence. It currently consists of 46 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.
1
14737: contig of 14737 bp in length
14738 14837: gap of 100 bp
14838 28083: contig of 13246 bp in length
28084 28183: gap of 100 bp
28184 40511: contig of 12328 bp in length
40512 40611: gap of 100 bp
40612 51888: contig of 11287 bp in length
51889 51998: gap of 100 bp
51999 64081: contig of 12093 bp in length
64082 64181: gap of 100 bp
64182 74827: contig of 10646 bp in length
74828 74927: gap of 100 bp
74928 84831: contig of 9904 bp in length
84832 84931: gap of 100 bp
84932 91900: contig of 6969 bp in length
91901 92000: gap of 100 bp
92001 99014: contig of 7014 bp in length
99015 99114: gap of 100 bp
99115 105728: contig of 6614 bp in length
105729 105828: gap of 100 bp
105829 110842: contig of 5014 bp in length
110843 110942: gap of 100 bp
110943 116457: contig of 5525 bp in length
116458 116567: gap of 100 bp
116568 121590: contig of 5023 bp in length
121591 121690: gap of 100 bp
121691 127550: contig of 5860 bp in length
127551 127650: gap of 100 bp
127651 132127: contig of 4477 bp in length
132128 132227: gap of 100 bp
132228 136727: contig of 4500 bp in length
136728 136827: gap of 100 bp
136828 141378: contig of 4551 bp in length
141379 141478: gap of 100 bp
141479 145693: contig of 4215 bp in length
145694 145793: gap of 100 bp
145794 149654: contig of 3861 bp in length
149655 149754: gap of 100 bp
149755 154794: contig of 5040 bp in length
154795 154894: gap of 100 bp
154895 158481: contig of 3587 bp in length
158482 158581: gap of 100 bp
158582 164776: contig of 6195 bp in length
164777 164876: gap of 100 bp
164877 167761: contig of 2885 bp in length
167762 167861: gap of 100 bp
167862 171113: contig of 3252 bp in length
171114 171213: gap of 100 bp
171214 174080: contig of 2867 bp in length
174081 174180: gap of 100 bp
174181 177056: contig of 2876 bp in length
177057 177156: gap of 100 bp
177157 179451: contig of 2295 bp in length
179452 179551: gap of 100 bp
179552 182040: contig of 2489 bp in length
182041 182140: gap of 100 bp
182141 184148: contig of 2278 bp in length
184149 184518: gap of 100 bp
184519 186523: contig of 2005 bp in length
186524 186623: gap of 100 bp
186624 188422: contig of 1839 bp in length
188423 188562: gap of 100 bp
188563 190196: contig of 1634 bp in length
190197 190296: gap of 100 bp
190297 192401: contig of 2105 bp in length
192402 192501: gap of 100 bp
192502 194597: contig of 2096 bp in length
194598 194697: gap of 100 bp
194698 196254: contig of 1557 bp in length
196255 196354: gap of 100 bp
196355 198593: contig of 2239 bp in length

* 138594 198693: gap of 100 bp
138694 200646: contig of 1953 bp in length
200647 200746: gap of 100 bp
200747 202409: contig of 1663 bp in length
202410 202509: gap of 100 bp
202510 204484: contig of 1975 bp in length
204485 204584: gap of 100 bp
204585 206037: contig of 1453 bp in length
206038 206137: gap of 100 bp
206138 207948: contig of 1811 bp in length
207949 208048: gap of 100 bp
208049 209677: contig of 1629 bp in length
209678 209777: gap of 100 bp
209778 210908: contig of 1131 bp in length
210909 211008: gap of 100 bp
211009 212394: contig of 1386 bp in length
212395 212494: gap of 100 bp
212495 213510: contig of 1016 bp in length
213511 213610: gap of 100 bp
213611 214701: contig of 1091 bp in length.
FEATURES
Location/Qualifiers
source
1..214701
/organism="Homo sapiens"
Query Match 71.5%; Score 18.6; DB 2; Length 214701;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 CAACGCGAGCGCCGATCTACGAA 26
DB 137150 CAACGCGAGCGCCGCTCTACAAA 137174
RESULT 28
AE003811
LOCUS
DEFINITION
Drosophila melanogaster chromosome 2R, section 39 of 74 of the complete sequence.
ACCESSION
AE003811 AE002787 AE013599
VERSION
AE003811.3 GI:21627138
KEYWORDS
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 255652)
AUTHORS
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazer V.R., Champs M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Cabor G.L., Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.V., Beres P.V., Bernier B.P., Bhandari D., Bolshakov S., Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W., Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K.,

Müsskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
 Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
 Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
 Weinstein,G.M., Weissbach,J., Williams,S.M., Woodruff,
 Worley,K.C., Wu,Q., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 10731132
 REFERENCE
 JOURNAL
 MEDLINE
 PUBMED
 AUTHORS
 2 (bases 1 to 255652)
 Celniker,S.E., Adams,M.D., Kronmiller,B., Wan,K.H., Holt,R.A.,
 Evans,C.A., Cocayne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y.,
 Banzon,J., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
 Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M.,
 Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
 Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
 Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
 Hegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
 Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
 Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
 Stapleton,M., Strong,R., Svirskas,R., Tector,C., Tyler,D.,
 Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.
 Sequencing of *Drosophila melanogaster* genome
 Unpublished
 3 (bases 1 to 255652)
 Misra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L.,
 Campbell,K., Hradscky,P., Huang,Y., Kaniker,J.S., Prochnik,S.E.,
 Smith,C.D., Tupy,J.L., Bergman,C.M., Berman,B.P., Carlson,J.W.,
 Celniker,S.E., Clump,M.E., Drysdale,R.A., Emmert,D., Frise,E., de
 Grey,A.D.N.J., Harris,N.L., Kronmiller,B., Marshall,B.,
 Milburn,G.H., Richter,J., Russo,S., Seale,S.M.J., Smith,E.,
 Shu,S., Smutniak,F., Whitfield,E.J., Ashburner,M., Gelbart,W.M.,
 Rubin,G.M., Mungall,C.J. and Lewis,S.E.
 Annotation of *Drosophila melanogaster* genome
 Unpublished
 4 (bases 1 to 255652)
 Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 5 (bases 1 to 255652)
 FlyBase
 Direct Submission
 Submitted (06-SEP-2002) University of California Berkeley, 539 Life
 Sciences Addition, Berkeley, CA 94720, USA
 6 (bases 1 to 255652)
 FlyBase
 Direct Submission
 Submitted (23-JAN-2003) University of California Berkeley, 539 Life
 Sciences Addition, Berkeley, CA 94720, USA
 On Jun 28, 2002 this sequence version replaced gi:10727530.
 Location/Qualifiers
 1. 255652
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /chromosome="2R"
 /note="genotype: y[1]; cn[1] bw[1] sp[1]; Rh6[1]"
 1884. .36739
 /gene="igl"
 /locus_tag="CG18285"
 /note="last curated on Tue Feb 26 11:36:14 PST 2002"
 /map="51E3-51E5"
 /db_xref="FLYBASE:FBgn0013467"

mRNA
 CDS
 mRNA
 CDS
 gene
 mRNA
 CDS
 gene
 mRNA
 CDS

join(1884. .2660,26958. .27155,35201. .36739)
 /gene="igl"
 /locus_tag="CG18285"
 /product="CG18285-RA"
 /db_xref="FLYBASE:FBgn0013467"
 join(2363. .2660,26958. .27155,35201. .35427)
 /gene="igl"
 /locus_tag="CG18285"
 /note="igl gene product from transcript CG18285-RA"
 /codon_start=1
 /product="CG18285-PA"
 /protein_id="AAF58168.2"
 /db_xref="GI:21627139"
 /db_xref="FLYBASE:FBgn0013467"
 /translation="MGCNTSQELTKDGAAMDVNSGEPFSPAPPLEGESKSSASNVH
 TNHAKSSIIISNGEAKAGCGAVGGSGKSEATNGIDRDCDKAAITFEFNDDEAKA
 ATKIQAVPRGHKVTMKKSETKTATNGSAGAAPSAASAAAEAEPTKALEAEF
 DPNDKDLCHAALKIQSTFRGHARLAKLYNKDAPEDEIQITKKVAELDIDLTPELN
 KAATKIQASFRGHTRKDAPE"
 join(11362. .11545,26958. .27155,35201. .36739)
 /gene="igl"
 /locus_tag="CG18285"
 /product="CG18285-RB"
 /db_xref="FLYBASE:FBgn0013467"
 join(27014. .27155,35201. .35427)
 /gene="igl"
 /locus_tag="CG18285"
 /note="igl gene product from transcript CG18285-RB"
 /codon_start=1
 /product="CG18285-PB"
 /protein_id="AAM68523.1"
 /db_xref="GI:21627140"
 /db_xref="FLYBASE:FBgn0013467"
 /translation="MKKSETKTATNGSAGAAPSAASAAAEAEPTKALEAEFP
 NDKDLCHAALKIQSTFRGHARLAKLYNKDAPEDEIQITKKVAELDIDLTPELNKA
 ATKIQASFRGHTRKDAPE"
 complement(45752. .47622)
 /locus_tag="CG8090"
 /note="last curated on Wed Feb 27 13:47:00 PST 2002"
 /map="51E5-51E5"
 /db_xref="FLYBASE:FBgn0033995"
 complement(join(45752. .46027,46080. .46265,46323. .46504,
 46564. .46748,46816. .47041,47102. .47372,47430. .47622))
 /locus_tag="CG8090"
 /product="CG8090-RA"
 /db_xref="FLYBASE:FBgn0033995"
 complement(join(45752. .46027,46080. .46265,46323. .46504,
 46564. .46748,46816. .47041,47102. .47372))
 /locus_tag="CG8090"
 /note="CG8090 gene product"
 /codon_start=1
 /product="CG8090-PA"
 /protein_id="AAF58167.1"
 /db_xref="GI:7303100"
 /db_xref="FLYBASE:FBgn0033995"
 /translation="MIFAGWLFNFKELFKHYEIRHISVOLIFSFSTFALSITWFELI
 IFEIDVLESSRYFHWRLGLTLLFTTAVIPIYICYSVTHISFSDKWRILTFIF
 MSYFIKLVNRIDGPFPLLSHAGHIFTIEQGVRSIVIGTVAILSGFAGVAPYPTS
 CWFTIKPVSRNDIICFERRALITVEMLSAKRKIAMAIYNNHKNLPSKPIWMLASA
 VORNTSGEDINOLKQVYVGFELLRSVLELSLKNMERQWSQTLKGYFNVLGH
 FFSYCYKIFMCCINIIFRVGRKDVTRGLEIAIHWCGNIDFAFNWQHISFLVGH
 CIVTSIRGLLLTLTKFIRISSKSSNIIVILGQINGMYFFSSVLLMRNNPAEYR
 VIITVEGLNLFHNFYHRWFDVIFLVSAITTIIVLYLSRKPKVRVDDSDLN"
 47969. .50918
 /locus_tag="CG11807"
 /note="last curated on Tue Feb 26 11:38:34 PST 2002"
 /map="51E5-51E5"
 /db_xref="FLYBASE:FBgn0033996"
 join(47969. .48276,48339. .48572,49467. .50346,50406. .50918)
 /locus_tag="CG11807"
 /note="last curated on Tue Feb 26 11:36:14 PST 2002"
 /db_xref="FLYBASE:FBgn0033996"
 join(48061. .48276,48339. .48572,49467. .50346,50406. .50551)

gene


```

/locus_tag="CG11807"
/feature="CG11807 gene product"
/codon_start=1
/product="CG11807-PA"
/protein_id="AAF38166.1"
/db_xref="GI:7303099"
/db_xref="FLYBASE:FBgn0033996"
/translation="MACYRQHADITTVTPKFSNESSGGVYTYDIDKRVKGVKWLVE
RYRDFANLHEKLVGEISISKKLLFPFKLVGNKOPFLQREQLIYLQELLIYFRT
ELPRALAEFLDFNKYDIYLLQDLAKLFNKGESDALLSKKEYNLSALEVAISERLSL
CPFSLDGGGDFSHVLDYCTQLVALVVTVPKDNASYAODYNTVDVPIRDSNIIPNR
LSPNLNAPRNLTKLFGSALSTENIVDIELKPTIQTICVNTITQNIQNLVLCNDLHFK
HCDVPSLLPTILASPGSGSPSTNGSALVSADAWQBITELDLTGNLLTQIDGVSRTA
PKLRLLLDQNRITVONLAELPOLQLLSLSCNLIACDVPHLTMGNLVLKLAQNKI
KTLGSLKLLSLVNLDSNNOIEIDENVHVNVLPLETURLTGNPLAGSVDRSRVL
ARFHERAAELSLDNEPQQLDALTALVLSALQSKQKQKQK"
Query Match 71.5%; Score 18.6; DB 3; Length 255652;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACACGGCGAGCCGCGAATCTACGA 25
Db 230535 AGAACGGCGACCGCGAATCTCCGA 230559

RESULT 29
AE017177/c
LOCUS AE017177 301618 bp DNA linear BCT 02-SEP-2003
DEFINITION Porphyromonas gingivalis W83 section 6 of 8 of the complete genome.
ACCESSION AE017177 AE015924
VERSION AE017177.1 GI:34397416
KEYWORDS
SOURCE
ORGANISM Porphyromonas gingivalis W83
Porphyromonas gingivalis W83
Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
Porphyromonadaceae; Porphyromonas.
REFERENCE 1 (bases 1 to 301618)
Daugherty, S., Dodson, R., Durkin, A., Gwinn, M., Haft, D., Kolonay, J.,
Nelson, W., White, O., Mason, T., Tallon, L., Gray, J., Granger, D.,
Tettelin, H., Dong, H., Galvin, J., Duncan, M., Dewhirst, F. and
Fraser, C.
Complete Genome Sequence of the Oral Pathogenic Bacterium
Porphyromonas gingivalis Strain W83
J. Bacteriol. 185 (18), 5591-5601 (2003)
REFERENCE 2 (bases 1 to 301618)
Nelson, K., Fleischmann, R., DeBoy, R., Paulsen, I., Fouts, D., Eisen, J.,
Daugherty, S., Dodson, R., Durkin, A., Gwinn, M., Haft, D., Kolonay, J.,
Nelson, W., White, O., Mason, T., Tallon, L., Gray, J., Granger, D.,
Tettelin, H., Dong, H., Galvin, J., Duncan, M., Dewhirst, F. and
Fraser, C.
Direct Submission
Submitted (29-OCT-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
FEATURES
source
location/Qualifiers
1. 301618
/organism="Porphyromonas gingivalis W83"
/mol_type="genomic DNA"
/strain="W83"
/db_xref="taxon:242619"
complement(203..361)
/locus_tag="PG1426"
complement(203..361)
/locus_tag="PG1426"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAQ66479.1"
/db_xref="GI:34397417"
/translation="MVSVERILMSIDRCVLLCVGWQDRGVGSVRAVKPIDKTVARH

PGDFTIVN"
complement(916..3447)
/locus_tag="PG1427"
complement(916..3447)
/locus_tag="PG1427"
/codon_start=1
/transl_table=11
/product="thiol protease/hemagglutinin PrtT precursor, putative"
/protein_id="AAQ66480.1"
/db_xref="GI:34397418"
/translation="MKKSELLAIVMLFGIAMOGHSPVTKERALSIALRLAQVSLRM
QOTAVSDKISIDVYVYRQGDARGITSQEESGFAIYVAVNRGNNEGTAUVAADRIPTI
LAYSPGIRGDMSPMDNRMWLOIYDQELGLSLGKAQNEELITLREGVPAVHALMD
NHFANDPRVWNGQYFNNKPELLPNHGHAYTCVATAAQIMRHYSPVLOQEGSFYD
HAGSLVNGWSGTGEGYWIWPNPDLNLTQSOVDATIMRDSVASVMSFYENG
SGTYSVYVVGALRNFRKRYKSLQALVRLYTSQEWHDIMRGLASGRPVYVAGNNQSI
GHAFVCDGVASDGTTFHNGWGVNGFVKLLTLLSPTSLIGIGEGIGFTIYQELITGI
EAKTPAEAGTADALPILAKLDJEARYKSEGLNAGVSYINTGEESQNLDLGLRLKAD
GEVSVKVTSSINISWYGEHPESFLAPNQLSQGINTILLIYRTTGTQWEPVRAHQ
GVVNSIKVNTDNNVVYVDNNEKLSIVNSFVADLNSYEHSITITVQFNSDSDPE
IRTPVAFALSTGATADDVISLGWMAEVFGSSNTFYVMKSDVLTLSGDYTLRYRFS
INNOKDEKKIGSVSKTPTETHTPLFEVGHNTQSTYTLDMAHNRVLPDFTLKNLGLP
FNGELVYVFPRTQSSGSLAAQETVHIKQGETFVYKPVVEQPIPGSVRATLHAFVN
GOOLYLVKGRNVTYKIVNGTAVEALESEEIRVFPNPARDVVEISAPCIPOETSIIL
FPLSGKIWMKNSLSAGHGRMDYSLRPNAGNYILLKVDGYTTKINIVH"
complement(3588..4073)
/genes="ribH"
/locus_tag="PG1428"
complement(3588..4073)
/genes="ribH"
/locus_tag="PG1428"
/notes="identified by match to TIGR protein family HMM
TIGR00114"
/codon_start=1
/transl_table=11
/product="riboflavin synthase, beta subunit"
/protein_id="AAQ66481.1"
/db_xref="GI:34397419"
/translation="WATAYHNLSYDYESVPCGKDLRTGIAVAENHNHITPELMKGA
I
DILLSHGVSADNIIIVHVPFTFELTYASAYLAQHEVDAIVAIGCVRGDTPHFHYIC
QVQTGVTQLNIDGFPVPVIFGVLTITETMLQAEERAGKGNKGTAAVTLAKMAGLER
I"
complement(4054..4164)
/locus_tag="PG1429"
complement(4054..4164)
/locus_tag="PG1429"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAQ66482.1"
/db_xref="GI:34397420"
/translation="MYQNDKKGIYMLGWMPPFTKLIROLQNHGYSLS"
complement(4165..4851)
/locus_tag="PG1430"
complement(4165..4851)
/locus_tag="PG1430"
/codon_start=1
/transl_table=11
/product="TPR domain protein"
/protein_id="AAQ66483.1"
/db_xref="GI:34397421"
/translation="MAKKENQPDNHGVSVSRSEQTEKNTKIGLVILGVITVAG
I
PAYKRFVSEPKAREAAKVAKYLAEDKFTOELDSALNGANGENGLLTVIKYSGSDA
SNLAKYAGICVYNLGEYQKAIIDLKGFSSKENVAPSIITRLIGDGYVQLKYEAVG
YFEKAAAASNDAITPGCLIKAGRYBELKQYDKXALAAKQVKKYYTSMANTVEAD
IIRVTKGAK"
complement(5037..5720)
/locus_tag="PG1431"
complement(5037..5720)
/locus_tag="PG1431"
/codon_start=1
```

/transl_table=11
/product="DNA-binding response regulator, LuxR family"
/protein_id="AA066484.1"
/db_xref="GI:34397422"
/translation="MTSIVLVDDHLYRVGLRAAISQMQLAIVGECSCSELETFPL
KEQPEGLIILDWLPDNGVNIARKLKLYPAVKIILISSEVSEKTVNELLAIDVVG
YICKAQINDGNAIRTVTGGSHFYKSVSMILDI CVRSADQSHKKKQOQOOND
SSBLTQREKVQVYLCDDGYSAKETAEMKMLSYRTVETHRSNHLHKLGFNAELIIRA
VKSGLVDWQ"
complement(5785..7650)
/locus_tag="PGI432"
complement(5785..7650)
/locus_tag="PGI432"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AA066485.1"
/db_xref="GI:34397423"
/translation="MFLTKNYCYCSROIQDTIMINIHTRLTILTLFLGCVFMLSKEIN
VDSLFQIADRTSATEKKRYEEMKQELHQGVGSAADAYSLAIKFCRDGGSTEEL
PFLIKYATWATHAGRYSNIAIMALEEIMRLIKNDNIILARTYMGIGVFFFOEKWD
DALVYERALQIAOKKEERGISAYNNMANIYQKNGTQOAYSYYNKALDIQRTGD
SVSCNSIMGTITMINEKNPDSYKPLKEALITASKINDIEIQAISYAHLAYVYAWR
SEYKANEVLTIAENLACHAGYNQTRALIDIASIYAEGVGLSLAIIESTLLVATITL
SIAKQOEKITEFDIRPKSKZAEIALKNQELHARKLHIALIISTLLVATITL
IIHAARYQRNKLRENDTNRLLSIIISHDIGPALAQKALDITLSSSDCGSWTR
NMLSSIRNSVATELTLLQNLDDWSNIQIGNLAPFVFNIVENIRVINEYSISARKN
DISLHLEAPKQCLVHADQOMINTVIRNLLSNAIKFSGAGSVIHRVVEEKEKAGIKVS
VSDGIGNSKKQITDILESNDKKKATAGTDEIGSLGILICKLLAKKKQVMH"
7932..8147
/locus_tag="PGI433"
7932..8147
/locus_tag="PGI433"
/note="identified by match to TIGR protein family HMM
TIGR01280"
/codon_start=1
/transl_table=11
/product="hydrolase"
/protein_id="AA066486.1"
/db_xref="GI:34397424"
/translation="MSENKTYTEAMRLEEVIRVIEHSPDVELTKLAEAIALIGF
CREKLTADVADKQIEELMAKLSWSKNAML"
8123..8791
/gene="ispb"
/locus_tag="PGI434"
8123..8791
/gene="ispb"
/note="similar to GB:M96326, SP:P20160, PID:179302, and
PID:28977; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="4-diphosphocytidylyl-2C-methyl-D-erythritol
synthase"
/protein_id="AA066487.1"
/db_xref="GI:34397425"
/translation="MEOKRVALLVAGGHGLRMGADPRKQFILLAGLPVLMLHNLNRFAP
HVDALVLPDTHAYWELCKRYDFSVSHRVWAGGNTRFASVRNGLVQVDPDGLVAV
HDGVRPLVSAETIDACPLAELKGAVPCRPMTESLRYATDGNVYVDRGRYVTQTP
QTRFSEWLEARQRYEYFTDDCSVYEZHFFGRPVALLVGNINIKLITPDLDSLAKL
LLTS"
9020..9096
/gene="tRNA-Asp-1"
9020..9096
/gene="tRNA-Asp-1"
/product="tRNA-Asp"
/note="codon recognized: GAC"
9404..10630
/locus_tag="PGI435"
9404..10630
/locus_tag="PGI435"
/codon_start=1
/transl_table=11

/product="integrase"
/protein_id="AA066488.1"
/db_xref="GI:34397426"
/translation="MKSQFRSVFYLRGNYINKEGKTPILIRIYNLKNRRLSLGSGTGLSV
NAQLDWSEKVKVGHSAEALVNRKIBIIRADILTIYKRLVTVDDITPERIKSEYCG
OTDLNLSVLEFDKNEVDRAQVINGINTAATLOKYENSKHFTFLKAKNRITDLKFS
ELTFLVHNFPIYLLTVAHCCNPATATKLLKFKITLYGRHGFSLDHPFVNRKPFHTQ
DSSGFLUDEBETQRIIGKISIPISPLDLIVRDLIFISCESGLAYIDVSNLTEDNIVELDG
REWINTNRQKTRVATNIIILLDIPRIKRIEYKAPVRKNGHLLPILSNQKNAYLKEIALD
71.5%; Score 18.6; DB 1; Length 301618;
Query Match 84.0%; Pred. No. le+03;
Best Local Similarity 84.0%; 0; Mismatches 4; Indels 0; Gaps 0;
Matches 21; Conservative
QY 2 CAACGGGAGGCCCGAATCTACGAA 26
Db 48065 CAACGGCATGCCCGGAATCAACGAA 48041
RESULT 30
BX569695 344615 bp DNA circular BCT 23-SEP-2003
LOCUS
Synecococcus sp. WH8102 complete genome; segment 7/7.
DEFINITION
ACCESSION BX569695 BX548020
VERSION BX569695.1 GI:33639324
KEYWORDS
complete genomes.
SOURCE
Synecococcus sp. WH 8102
ORGANISM
Synecococcus sp. WH 8102
REFERENCE
1 (bases 1 to 344615)
AUTHORS
Palenik,B., Brahamsa,B., Larimer,F., Land,M., Hauser,L., Chain,P.,
Lamerdin,J., Regala,W., Allen,E.A., McCarren,J., Paulsen,I.,
Dufresne,A., Partensky,F., Webb,E. and Waterbury,J.
The genome of a motile marine Synecococcus
Nature 424 (6952), 1037-1042 (2003)
22825697
12917641
2 (bases 1 to 344615)
Larimer,F. and Palenik,B.
Synecococcus genome consortium
Direct Submission
Submitted (01-JUL-2003) Larimer, F., Submitted on behalf of the
Synecococcus genome consortium, the DOE Joint Genome Institute, CA
Production Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA
94598, USA, and the Genome Analysis Group, Oak Ridge National
Laboratory, 1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
larimerf@ornl.gov
Location/Qualifiers
1..344615 Synecococcus sp. WH 8102"
/organism="Synecococcus sp. WH 8102"
/mol_type="genomic DNA"
/strain="WH8102"
/db_xref="taxon:84588"
67..1716
/gene="nadB"
/locus_tag="SYNW2192"
67..1716
/gene="nadB"
/locus_tag="SYNW2192"
/EC_number="1.4.3.16"
/note="putative assignment"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative L-aspartate oxidase"
/protein_id="CAE08707.1"
/db_xref="GI:33639325"
/translation="MDPSRPLDAIDPGWVWVVGAGAGLMTCLLEPEGLSVLLNLR
NTRGRSSRWAGGIAAVTRPDSAGSHADTLNAGAGLCDDGDAVRFVLDAAAPCVDR
LQGLGNAPDGDGHGLATTLEAAHSHRRVHVODTRGALVDLVRVRERRPGLHRRG
VNRVLVDVQNGCGGVQVLDGSHLOHIOARAVVLASGGGHLFANTTNPAQACGGGIA
LAWAGAAVEDELFVQFHFTALRLDAPCFILSEAVRGEGVLDVDAQSGSPVDHLLQR
DLAPRDQVSRAIVQRWEGVQVQMWLDFSAIPRAQNAERPFILDRCSFEGNPLERP

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 149.397 Seconds
(without alignments)
739.327 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26

Sequence: 1 acaacggcagggccgaatctacgaa 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

N Geneseqn_29Jan04:.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2003as.*
8: Geneseqn2003bs.*
9: Geneseqn2003cs.*
10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26	100.0	393	2	AAT49174
C 2	26	100.0	110000	4	AAI99682_42
C 3	26	100.0	110000	4	AAI99683_42
C 4	18.6	71.5	2623	4	ABL11529
C 5	18.6	71.5	6747	4	ABL11528
C 6	18.6	71.5	14705	4	AAAS9523
C 7	18.6	71.5	14705	7	ACF64452
C 8	18.2	70.0	546	6	ABQ90865
C 9	18	69.2	656	6	ABQ18285
C 10	18	69.2	656	6	ABQ18284
C 11	18	69.2	1113	6	ABQ46934
C 12	18	69.2	1113	6	ABQ46935
C 13	18	69.2	2050	6	ABQ39756
C 14	18	69.2	2050	6	ABQ39757
C 15	18	69.2	110000	4	AAI99682_04
C 16	18	69.2	110000	4	AAI99682_05
C 17	18	69.2	110000	4	AAI99683_04
C 18	18	69.2	110000	4	AAI99683_05
C 19	17.6	67.7	839	6	ABQ16680
C 20	17.6	67.7	839	6	ABQ16681
C 21	17.6	67.7	851	6	ABQ33932
C 22	17.6	67.7	851	6	ABQ33933
C 23	17.6	67.7	1017	7	ACA24162

24	17.6	67.7	11340	4	ABL16852	AbL16852 Drosophil
C 25	17.6	67.7	110000	2	AAV30458_0	AAV30458 Rhizobium
C 26	17.6	67.7	110000	2	AAV30459_0	AAV30459 Rhizobium
C 27	17.2	66.2	34980	6	ABQ1844	AbQ1844 Bifidobac
C 28	17	65.4	530	6	ABQ14535	AbQ14535 Oligonuc
C 29	17	65.4	530	6	ABQ14534	AbQ14534 Oligonuc
C 30	17	65.4	879	3	AAAS5205	AAAS5205 C. symbio
C 31	17	65.4	913	3	AAAF08967	AAAF08967 Fusarium
C 32	17	65.4	934	6	ABQ45717	AbQ45717 Oligonuc
C 33	17	65.4	934	6	ABQ45716	AbQ45716 Oligonuc
C 34	17	65.4	1173	7	ACA25903	ACA25903 Prokaryot
C 35	17	65.4	1225	4	ABL08271	AbL08271 Drosophil
C 36	17	65.4	4568	4	ABL08270	AbL08270 Drosophil
C 37	17	65.4	4432	3	AAAS5187	AAAS5187 Cenarchae
C 38	17	65.4	93483	8	ADA03083	ADA03083 Mouse mCG
C 39	17	65.4	93483	8	ADA66367	ADA66367 Mouse mCG
C 40	17	65.4	93483	9	ADB72821	ADB72821 Mouse mCG
C 41	17	65.4	34980	6	ABQ18447	AbQ18447 Bifidobac
C 42	16.8	64.6	637	7	ABZ533960	ABZ533960 Aspergill
C 43	16.8	64.6	1230	1	AAAN70053	AAAN70053 Benzene o
C 44	16.8	64.6	1230	1	AAAN70677	AAAN70677 Sequence
C 45	16.8	64.6	1230	1	AAAN70822	AAAN70822 DNA encod
C 46	16.8	64.6	1755	7	ADA70624	ADA70624 Rice gene
C 47	16.8	64.6	5274	1	AAAN70679	AAAN70679 Sequence
C 48	16.8	64.6	5274	1	AAAN70824	AAAN70824 Catechol
C 49	16.6	63.8	290	6	ABL73770	AbL73770 Corn tass
C 50	16.6	63.8	932	6	ABQ34521	AbQ34521 Oligonuc
C 51	16.6	63.8	932	6	ABQ34920	AbQ34920 Oligonuc
C 52	16.6	63.8	1431	4	ABL08403	AbL08403 Drosophil
C 53	16.6	63.8	1455	7	ACA30126	ACA30126 Prokaryot
C 54	16.6	63.8	2031	9	ADD46116	ADD46116 Rat gene
C 55	16.6	63.8	2031	9	ABE61715	ABE61715 Rat gene
C 56	16.6	63.8	2157	7	ACA27237	ACA27237 Prokaryot
C 57	16.6	63.8	2307	4	ABL23687	AbL23687 Drosophil
C 58	16.6	63.8	3077	7	ABZ10070	ABZ10070 Haematopo
C 59	16.6	63.8	4218	9	ADC87164	ADC87164 Human GFC
C 60	16.6	63.8	5479	4	ABL23686	AbL23686 Drosophil
C 61	16.6	63.8	6077	6	ABL33855	AbL33855 Human mm
C 62	16.6	63.8	6077	9	ADB54126	ADB54126 Pretreate
C 63	16.6	63.8	6077	9	ADB84126	ADB84126 Human lym
C 64	16.6	63.8	6077	9	ABL08402	AbL08402 Drosophil
C 65	16.6	63.8	7543	8	AAAL61505	AAAL61505 Human inh
C 66	16.6	63.8	27425	3	AAAS3978	AAAS3978 DMP3 tum
C 67	16.6	63.8	34769	4	AAAS46775	AAAS46775 Tumour su
C 68	16.6	63.8	67832	8	ADA02801	ADA02801 Mouse itk
C 69	16.6	63.8	67832	9	ADB72539	ADB72539 Mouse itk
C 70	16.6	63.8	67832	9	ADC85281	ADC85281 Mouse itk
C 71	16.6	63.8	156638	6	ABQ81850	AbQ81850 Bifidobac
C 72	16.4	63.1	511	6	ABQ50581	AbQ50581 Oligonuc
C 73	16.4	63.1	511	6	ABQ50580	AbQ50580 Oligonuc
C 74	16.4	63.1	522	6	ABQ33181	AbQ33181 Oligonuc
C 75	16.4	63.1	522	6	ABQ33180	AbQ33180 Oligonuc
C 76	16.4	63.1	566	6	ABQ50784	AbQ50784 Oligonuc
C 77	16.4	63.1	566	6	ABQ50785	AbQ50785 Oligonuc
C 78	16.4	63.1	595	3	AAAC38541	AAAC38541 Arabidops
C 79	16.4	63.1	613	6	ABQ29765	AbQ29765 Oligonuc
C 80	16.4	63.1	613	6	ABQ29764	AbQ29764 Oligonuc
C 81	16.4	63.1	643	6	ABQ28009	AbQ28009 Oligonuc
C 82	16.4	63.1	643	6	ABQ28008	AbQ28008 Oligonuc
C 83	16.4	63.1	699	7	ACA42767	ACA42767 Prokaryot
C 84	16.4	63.1	717	6	ABQ23027	AbQ23027 Oligonuc
C 85	16.4	63.1	717	6	ABQ23026	AbQ23026 Oligonuc
C 86	16.4	63.1	723	6	ABQ40140	AbQ40140 Oligonuc
C 87	16.4	63.1	723	6	ABQ40141	AbQ40141 Oligonuc
C 88	16.4	63.1	734	6	ABQ38865	AbQ38865 Oligonuc
C 89	16.4	63.1	734	6	ABQ38864	AbQ38864 Oligonuc
C 90	16.4	63.1	748	6	ABQ24376	AbQ24376 Oligonuc
C 91	16.4	63.1	748	6	ABQ24377	AbQ24377 Oligonuc
C 92	16.4	63.1	798	6	ABQ29802	AbQ29802 Oligonuc
C 93	16.4	63.1	798	6	ABQ29803	AbQ29803 Oligonuc
C 94	16.4	63.1	836	6	ABQ13596	AbQ13596 Oligonuc
C 95	16.4	63.1	836	6	ABQ13597	AbQ13597 Oligonuc
C 96	16.4	63.1	846	6	ABQ46632	AbQ46632 Oligonuc

97	16.4	63.1	846	6	ABQ46633	Abq46633 Oligonuc1	170	16	61.5	754	6	ABQ34323	Abq34323 Oligonuc1
98	16.4	63.1	874	6	ABQ44403	Abq44403 Oligonuc1	171	16	61.5	782	6	ABQ53104	Abq53104 Oligonuc1
99	16.4	63.1	876	6	ABQ44402	Abq44402 Oligonuc1	172	16	61.5	782	6	ABQ53105	Abq53105 Oligonuc1
100	16.4	63.1	877	6	ABQ23264	Abq23264 Oligonuc1	173	16	61.5	897	7	ADA70243	Ada70243 Rice gene
101	16.4	63.1	877	6	ABQ23265	Abq23265 Oligonuc1	174	16	61.5	897	8	ADA48201	Ada48201 Rice gene
102	16.4	63.1	876	6	ABQ52966	Abq52966 Oligonuc1	175	16	61.5	941	3	ACQ95102	Acq95102 Cat flea
103	16.4	63.1	976	6	ABQ52967	Abq52967 Oligonuc1	176	16	61.5	969	7	ACA25173	AcA25173 Prokaryot
104	16.4	63.1	976	6	ABQ52967	Abq52967 Oligonuc1	177	16	61.5	969	7	ACA25173	AcA25173 Prokaryot
105	16.4	63.1	998	6	ABQ33244	Abq33244 Oligonuc1	178	16	61.5	984	6	ABQ27350	Abq27350 Human clo
106	16.4	63.1	998	6	ABQ33245	Abq33245 Oligonuc1	179	16	61.5	984	6	ABQ27351	Abq27351 Oligonuc1
107	16.4	63.1	1204	6	ABQ14447	Abq14447 Oligonuc1	180	16	61.5	984	6	ABQ27351	Abq27351 Oligonuc1
108	16.4	63.1	1240	6	ABQ41114	Abq41114 Oligonuc1	181	16	61.5	984	6	ABQ27351	Abq27351 Oligonuc1
109	16.4	63.1	1240	6	ABQ41115	Abq41115 Oligonuc1	182	16	61.5	1080	7	ACA43613	AcA43613 Prokaryot
110	16.4	63.1	1312	6	ABQ30092	Abq30092 Oligonuc1	183	16	61.5	1227	6	ABK73883	Abk73883 Bacillus
111	16.4	63.1	1312	6	ABQ30093	Abq30093 Oligonuc1	184	16	61.5	1242	4	AA54493	Aa54493 Botulinum
112	16.4	63.1	1422	7	ACA38029	AcA38029 Prokaryot	185	16	61.5	1457	4	AAQ92642	AaQ92642 Human Nck
113	16.4	63.1	1422	7	ACA38029	AcA38029 Prokaryot	186	16	61.5	1457	4	AAQ92642	AaQ92642 Human Nck
114	16.4	63.1	1579	6	ABQ35415	Abq35415 Oligonuc1	187	16	61.5	1737	6	ABA05826	AbA05826 Arthrobac
115	16.4	63.1	1681	6	ABQ33738	Abq33738 Oligonuc1	188	16	61.5	1737	6	ABA05826	AbA05826 Arthrobac
116	16.4	63.1	1681	6	ABQ33738	Abq33738 Oligonuc1	189	16	61.5	1761	4	ABL18843	AbL18843 Drosophil
117	16.4	63.1	1726	5	AA572093	Aa572093 DNA encod	190	16	61.5	1884	6	ABA05824	AbA05824 Arthrobac
118	16.4	63.1	2000	6	ABZ17079	Abz17079 Arabidops	191	16	61.5	1956	5	AA536770	Aa536770 DNA encod
119	16.4	63.1	2265	7	AAH65400	Aah65400 C glutami	192	16	61.5	1985	2	AA536770	Aa536770 DNA encod
120	16.4	63.1	2265	7	AAH65400	Aah65400 C glutami	193	16	61.5	1985	2	AA536770	Aa536770 DNA encod
121	16.4	63.1	2670	7	ABT17795	Abt17795 Aspergill	194	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
122	16.4	63.1	2837	7	ABT17795	Abt17795 Aspergill	195	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
123	16.4	63.1	3499	5	AA580109	Aa580109 Human nov	196	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
124	16.4	63.1	3499	5	AA580109	Aa580109 Human nov	197	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
125	16.4	63.1	3499	5	AA580109	Aa580109 Human nov	198	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
126	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	199	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
127	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	200	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
128	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	201	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
129	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	202	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
130	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	203	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
131	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	204	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
132	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	205	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
133	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	206	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
134	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	207	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
135	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	208	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
136	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	209	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
137	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	210	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
138	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	211	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
139	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	212	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
140	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	213	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
141	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	214	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
142	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	215	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
143	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	216	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
144	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	217	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
145	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	218	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
146	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	219	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
147	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	220	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
148	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	221	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
149	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	222	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
150	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	223	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
151	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	224	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
152	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	225	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
153	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	226	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
154	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	227	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
155	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	228	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
156	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	229	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
157	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	230	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
158	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	231	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
159	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	232	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
160	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	233	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
161	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	234	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
162	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	235	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
163	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	236	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
164	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	237	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
165	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	238	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
166	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	239	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
167	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	240	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
168	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	241	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1
169	16.4	63.1	3969	5	AA580109	Aa580109 Human nov	242	16	61.5	2157	6	ABQ30957	Abq30957 Oligonuc1


```

WP AAI99682_15 1500001 1610000
WP AAI99682_16 1600001 1710000
WP AAI99682_17 1700001 1810000
WP AAI99682_18 1800001 1910000
WP AAI99682_19 1900001 2010000
WP AAI99682_20 2000001 2110000
WP AAI99682_21 2100001 2210000
WP AAI99682_22 2200001 2310000
WP AAI99682_23 2300001 2410000
WP AAI99682_24 2400001 2510000
WP AAI99682_25 2500001 2610000
WP AAI99682_26 2600001 2710000
WP AAI99682_27 2700001 2810000
WP AAI99682_28 2800001 2910000
WP AAI99682_29 2900001 3010000
WP AAI99682_30 3000001 3110000
WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529

Query Match 100.0%; Score 26; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACAACGGCGAGCCGCAATCTACGAA 26
Db 26924 ACAACGGCGAGCCGCAATCTACGAA 26949

RESULT 3
Continuation (43 of 44) of AAI99683 from base 4200001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000

Query Match 100.0%; Score 26; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACAACGGCGAGCCGCAATCTACGAA 26
Db 19170 ACAACGGCGAGCCGCAATCTACGAA 19195

RESULT 4
ABLI1529
ID ABLI1529 standard; cDNA; 2623 BP.
XX
XX ABLI1529;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 29059.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEXE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW,
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB67426.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 29069; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
XX sequences (ABLI01840-ABLI16175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pat_sequences

```


XX SQ Sequence 2623 BP; 741 A; 673 C; 694 G; 515 T; 0 U; 0 Other;
 XX SQ Query Match 71.5%; Score 18.6; DB 4; Length 2623;
 XX Best Local Similarity 84.0%; Pred. No. 53;
 XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 ACAACGGCGAGCGCCGAATCTACGA 25
 DB 1566 AGAACGGCGAGCGCCGAATCTCCGA 1590
 RESULT 5
 ABL11528/c
 ID ABL11528 standard; cDNA; 6747 BP.
 XX AC ABL11528;
 XX DT 26-MAR-2002 (first entry)
 XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 29066.
 XX KW Drosophila; developmental biology; cell signalling; insecticide;
 XX KW pharmaceutical; gene; ss.
 XX OS Drosophila melanogaster.
 XX WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US009231.
 XX PR 23-MAR-2000; 2000US-0191637P.
 XX PR 11-JUL-2000; 2000US-00614150.
 XX PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 WPI; 2001-656860/75.
 P-PSDB; ABB67425.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 PS Claim 1; SEQ ID NO 29066; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 6747 BP; 1672 A; 1591 C; 1660 G; 1824 T; 0 U; 0 Other;
 XX SQ Query Match 71.5%; Score 18.6; DB 4; Length 6747;
 XX Best Local Similarity 84.0%; Pred. No. 58;
 XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 ACAACGGCGAGCGCCGAATCTACGA 25
 DB 2375 AGAACGGCGAGCGCCGAATCTCCGA 2351
 RESULT 6
 AAS59523

ID AAS59523 standard; DNA; 14705 BP.
 XX AAS59523;
 XX DT 13-FEB-2002 (first entry)
 XX DE Propionibacterium acnes immunogenic protein encoding DNA #18.
 XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant; ds.
 XX OS Propionibacterium acnes.
 XX WO200181581-A2.
 XX PD 01-NOV-2001.
 XX PF 20-APR-2001; 2001WO-US012865.
 XX PR 21-APR-2000; 2000US-0199047P.
 XX PR 02-JUN-2000; 2000US-0208841P.
 XX PR 07-JUL-2000; 2000US-0216747P.
 XX PA (CORI-) CORIXA CORP.
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 WPI; 2001-616774/71.
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 XX vaccinating against and diagnosing infections, especially useful for
 XX treating acne vulgaris.
 PS Claim 1; SEQ ID NO 18; 1069pp; English.
 CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
 CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
 CC associated DNA sequences are used in the treatment, prevention and
 CC diagnosis of medical conditions caused by P. acnes. The disorders include
 CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in
 CC infections of bone, joints and the central nervous system, however it is
 CC particularly involved in the inflammatory lesions associated with acne
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a
 CC patient comprises contacting a sample with a binding agent that binds to
 CC the proteins of the invention and determining the amount of bound protein
 CC in the sample. The polypeptides may be used as antigens in the production
 CC of antibodies specific for P. acnes proteins. These antibodies can be
 CC used to downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
 CC polypeptides shown in AAU44569-AAU44720 and AAU67491-AAU67493. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 14705 BP; 3295 A; 4621 C; 4092 G; 2692 T; 0 U; 5 Other;
 XX SQ Query Match 71.5%; Score 18.6; DB 4; Length 14705;
 XX Best Local Similarity 84.0%; Pred. No. 61;
 XX Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 ACAACGGCGAGCGCCGAATCTACGA 25
 DB 4255 ACAACGTCGTCGCCGCACTACGA 4279
 RESULT 7
 ACF64452

```

ID ACP64452 standard; DNA; 14705 BP.
XX
AC ACP64452;
XX
DT 17-OCT-2003 (first entry)
XX
DE Propionibacterium acnes DNA contig sequence #18.
XX
KW Acne vulgaris; antisecborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine; ds.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallie-Douglas J;
XX
DR WPI; 2003-381789/36.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Claim 1; SEQ ID NO 18; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a P. acnes DNA contig which is specifically claimed
CC in the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 14705 BP; 3295 A; 4621 C; 4092 G; 2692 T; 0 U; 5 Other;

Query Match 71.5%; Score 18.6; DB 7; Length 14705;
Best Local Similarity 84.0%; Pred. No. 61;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCCGCAATCTACGA 25
Db 4255 ACAACGTCGTCGCCCGCAATCTACGA 4279

ACPF64452 standard; DNA; 14705 BP.
XX
AC ACP64452;
XX
DT 17-OCT-2003 (first entry)
XX
DE Propionibacterium acnes DNA contig sequence #18.
XX
KW Acne vulgaris; antisecborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine; ds.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallie-Douglas J;
XX
DR WPI; 2003-381789/36.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
PS Claim 1; SEQ ID NO 18; 1481pp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a P. acnes DNA contig which is specifically claimed
CC in the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 14705 BP; 3295 A; 4621 C; 4092 G; 2692 T; 0 U; 5 Other;

Query Match 71.5%; Score 18.6; DB 7; Length 14705;
Best Local Similarity 84.0%; Pred. No. 61;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCCGCAATCTACGA 25
Db 4255 ACAACGTCGTCGCCCGCAATCTACGA 4279

RESULT 8
ABQ0865
ID ABQ0865 standard; DNA; 546 BP.
XX
AC ABQ0865;
XX
DT 01-OCT-2002 (first entry)
XX
DE M. capsulatus gene #850 for DNA array.
XX
KW Micro array; gene; ds; differential expression; gene expression.
XX
OS Methylococcus capsulatus.
XX
PN WO200255655-A2.
XX
PD 18-JUL-2002.
XX
PF 14-JAN-2002; 2002WO-NO000019.
XX
PR 12-JAN-2001; 2001NO-00000235.
PR 12-JAN-2001; 2001NO-00000239.
XX
PA (UNIF-) UNIFOB STIFTTELSEN UNIV BERGEN.
PA (TIGR-) TIGR.
XX
PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
PI Salzberg SL;
XX
DR WPI; 2002-557818/59.
XX
PT Novel DNA array useful for determining differential expression of
PT Methylococcus capsulatus genes, comprises polynucleotides or
PT oligonucleotides representative for a selective number of Methylococcus
PT capsulatus genes.
XX
PS Claim 14; Page 374; 678pp; English.
XX
CC The invention relates to a novel DNA array giving a representation of a
CC number of Methylococcus capsulatus genes. The method of the invention is
CC useful for determination of the differential expression of the genes of
CC M. capsulatus, and for studying gene expression on a genomic scale and in
CC gene expression assays of M. capsulatus genes. The sequences shown in
CC ABQ00016-ABQ01855 represent M. capsulatus genes for use in arrays of the
CC invention
XX
SQ Sequence 546 BP; 100 A; 208 C; 164 G; 74 T; 0 U; 0 Other;

Query Match 70.0%; Score 18.2; DB 6; Length 546;
Best Local Similarity 87.0%; Pred. No. 73;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCCGCAATCTACG 24
Db 126 CAACGGCGAGGCCCGCAATCTACG 148

RESULT 9
ABQ18285
ID ABQ18285 standard; DNA; 656 BP.
XX
AC ABQ18285;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 4876.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
```


PF 01-SEP-2001; 2001WO-EP010074.
XX
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
XX
PA (EPIG-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC status of many C residues to be determined simultaneously. ABQ13410-
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1113 BP; 162 A; 140 C; 418 G; 393 T; 0 U; 0 Other;
Query Match 69.2%; Score 18; DB 6; Length 1113;
Best Local Similarity 80.8%; Pred. No. 96;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ACAACGGCGAGCCCGAATCTACGAA 26
Db 765 AAAACGACGACGACCGAATCTACGAA 740
RESULT 12
ABQ46935
ID ABQ46935 standard; DNA; 1113 BP.
XX
XX AC ABQ46935;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 33526.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX
XX PR 01-SEP-2000; 2000DE-01043826.
XX
XX PR 01-SEP-2000; 2000DE-01043826.
XX
XX PA (EPIG-) EPIGENOMICS AG.

PR 05-SEP-2000; 2000DE-01044543.
XX
XX
PA (EPIG-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX WPI; 2002-371829/40.
XX
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC status of many C residues to be determined simultaneously. ABQ13410-
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 1113 BP; 393 A; 418 C; 140 G; 162 T; 0 U; 0 Other;
Query Match 69.2%; Score 18; DB 6; Length 1113;
Best Local Similarity 80.8%; Pred. No. 96;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ACAACGGCGAGCCCGAATCTACGAA 26
Db 349 AAAACGACGACGACCGAATCTACGAA 374
RESULT 13
ABQ39756/c
ID ABQ39756 standard; DNA; 2050 BP.
XX
XX AC ABQ39756;
XX
XX DT 12-JUL-2002 (first entry)
XX
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26347.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP010074.
XX
XX PR 01-SEP-2000; 2000DE-01043826.
XX
XX PR 05-SEP-2000; 2000DE-01044543.
XX
XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 XX DR WPI; 2002-371829/40.
 XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
 XX PT diagnosis and prognosis, comprises selective hybridization of amplicons
 XX PT from chemically treated DNA.
 XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX SQ Sequence 2050 BP; 241 A; 276 C; 766 G; 767 T; 0 U; 0 Other;
 Query Match 69.2%; Score 18; DB 6; Length 2050;
 Best Local Similarity 80.8%; Pred. No. 1e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 ACAACGGCGAGCGCCGAATCTACGAA 26
 DB 1203 ATACCGCGACGACCGAACTACGAA 1178
 RESULT 14
 ABQ39757
 ID ABQ39757 standard; DNA; 2050 BP.
 AC ABQ39757;
 XX DT 12-JUL-2002 (first entry)
 XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26348.
 XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 XX KW drug; side effect; cancer; central nervous system; cardiovascular;
 XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX KW SNP; cell differentiation; ds.
 XX OS Homo sapiens.
 XX FN WO200218632-A2.
 XX PD 07-MAR-2002.
 XX PF 01-SEP-2001; 2001WO-EP010074.
 XX PR 01-SEP-2000; 2000DE-01043826.
 XX PR 05-SEP-2000; 2000DE-01044543.
 XX XX (EPIG-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.
 XX PT Determining the degree of cytosine methylation in genomic DNA, useful for
 XX PT diagnosis and prognosis, comprises selective hybridization of amplicons
 XX PT from chemically treated DNA.
 XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one member,
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
 CC degree of hybridisation to both classes is determined from the label on
 CC the amplicon. From the ratio of labels hybridised to the two classes of
 CC oligomers, the degree of methylation is calculated. The method is used:
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
 CC and of a wide range of diseases, e.g. cancer, disorders of the central
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
 CC particularly by detecting mutations or single nucleotide polymorphisms
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for
 CC investigating cell differentiation. The method allows the methylation
 CC status of many C residues to be determined simultaneously. ABQ13410-
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method
 CC for determining the degree of cytosine methylation described in the
 CC disclosure of the invention
 XX SQ Sequence 2050 BP; 767 A; 766 C; 276 G; 241 T; 0 U; 0 Other;
 Query Match 69.2%; Score 18; DB 6; Length 2050;
 Best Local Similarity 80.8%; Pred. No. 1e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 ACAACGGCGAGCGCCGAATCTACGAA 26
 DB 848 ATACCGCGACGACCGAACTACGAA 873
 RESULT 15
 AA199682_04
 Continuation (5 of 45) of AA199682 from base 400001 (Mycobacterium tuberculosis strain H
 WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682
 WP Fragment Name Begin End
 WP AA199682_00 1 113000
 WP AA199682_01 100001 210000
 WP AA199682_02 200001 310000
 WP AA199682_03 300001 410000
 WP AA199682_04 400001 510000
 WP AA199682_05 500001 610000
 WP AA199682_06 600001 710000
 WP AA199682_07 700001 810000
 WP AA199682_08 800001 910000
 WP AA199682_09 900001 1010000
 WP AA199682_10 1000001 1110000
 WP AA199682_11 1100001 1210000
 WP AA199682_12 1200001 1310000
 WP AA199682_13 1300001 1410000
 WP AA199682_14 1400001 1510000
 WP AA199682_15 1500001 1610000
 WP AA199682_16 1600001 1710000
 WP AA199682_17 1700001 1810000
 WP AA199682_18 1800001 1910000
 WP AA199682_19 1900001 2010000
 WP AA199682_20 2000001 2110000
 WP AA199682_21 2100001 2210000
 WP AA199682_22 2200001 2310000
 WP AA199682_23 2300001 2410000
 WP AA199682_24 2400001 2510000
 WP AA199682_25 2500001 2610000
 WP AA199682_26 2600001 2710000
 WP AA199682_27 2700001 2810000

RESULT 18

AAI99683_05
Continuation (6 of 44) of AAI99683 from base 500001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End

WP AAI99683_00 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Query Match 69.2%; Score 18; DB 4; Length 110000;
Best Local Similarity 80.8%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACACGGCGAGCGCCGAATCTACGAA 26
Db 7446 AGACCGCGAGGTGCGAATCAACGAA 7471

RESULT 19

ABQ16680/c
ID ABQ16680 standard; DNA; 839 BP.

XX AC
XX ABQ16680;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 3271.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP010074.

XX 01-SEP-2000; 2000DE-01043826.

XX 05-SEP-2000; 2000DE-01044543.

XX (EPIT-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX diagnosis and prognosis, comprises selective hybridization of amplicons
XX from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one member,
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
XX degree of hybridisation to both classes is determined from the label on
XX the amplicon. From the ratio of labels hybridised to the two classes of
XX oligomers, the degree of methylation is calculated. The method is used:
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
XX and of a wide range of diseases, e.g. cancer, disorders of the central
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
XX particularly by detecting mutations or single nucleotide polymorphisms
XX (SNP's); and (ii) for differentiation of cell or tissue types and for
XX investigating cell differentiation. The method allows the methylation
XX status of many C residues to be determined simultaneously. ABQ13410-
XX ABQ54121 represent genomic DNA sequences used to illustrate the method
XX for determining the degree of cytosine methylation described in the
XX disclosure of the invention

XX Sequence 839 BP; 96 A; 89 C; 284 G; 370 T; 0 U; 0 Other;

Query Match 67.7%; Score 17.6; DB 6; Length 839;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 AACGCGAGCGCCGAATCTACGAA 26

Db 458 AACGACGACGCGCGAATCTAAAAA 435

RESULT 20

ABQ16681

ID ABQ16681 standard; DNA; 839 BP.

XX AC
XX ABQ16681;

DT 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 3272.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.

XX OS Homo sapiens.

PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations of single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. AB013410-
CC ABO54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX Sequence 851 BP; 263 A; 357 C; 112 G; 119 T; 0 U; 0 Other;

Query Match 67.7%; Score 17.6; DB 6; Length 851;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ACAACGCGAGCGCCGAATCTACG 24
Db 505 AAAAAAGACGACGCGCCGAATCTACG 528

RESULT 23
ACA24162
ID ACA24162 standard; DNA; 1017 BP.
XX
XX ACA24162;
XX
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #5819.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Borrelia cepacia.
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX P-PSDB; ABU20292.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 12032; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1017 BP; 184 A; 334 C; 328 G; 171 T; 0 U; 0 Other;

Query Match 67.7%; Score 17.6; DB 7; Length 1017;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 CAACGCGAGCGCCGAATCTACGA 25
Db 798 CATCGCGAGCGCCCAATCTACGA 821

RESULT 24
ABL16852
ID ABL16852 standard; DNA; 11340 BP.
XX
XX ABL16852;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 2029.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX

XX DR WPI; 1998-110606/10.

XX PT New isolated symbiotic plasmid from *Rhizobium* sp. NGR234 - used to

XX PT develop products for modifying plant characteristics, e.g. nitrogen

XX PT fixation, synthesis of compounds and stress response.

XX PS Claim 1; Fig 3; 228pp; English.

XX CC This is the nucleotide sequence of the plasmid pNGR234a isolated from

XX CC *Rhizobium* sp. NGR234. Open reading frames (ORF) derivable from the

XX CC nucleotide sequence are claimed. The nucleotide sequences or ORFs can be

XX CC used e.g. in the transportation of compounds to and from an organism

XX CC which is a host to at least one of the nucleotide sequences, ORFs or

XX CC proteins, the degradation and/or metabolism of organic, inorganic,

XX CC natural or xenobiotic substances in a host organism or the modification

XX CC of the host range, nitrogen fixation abilities, for obtaining a synthetic

XX CC minimal set of ORFs required for functional *Rhizobium-legume* symbiosis,

XX CC especially for nodulation efficiency on host plants

XX SQ Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 U; 0 Other;

Query Match 67.7%; Score 17.6; DB 2; Length 110000;

Best Local Similarity 83.3%; Pred. No. 2.1e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGCGAGGCGCGCAATCTACGAA 26

DB 85320 AAGGCGCGCGCGCAACACGAA 85297

RESULT 27

ABQ81844

ID ABQ81844 standard; DNA; 349980 BP.

XX AC ABQ81844;

XX DT 19-NOV-2002 (first entry)

XX DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1100.

XX KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;

XX KW antidiarrheic; antibacterial; inhibitor of *Salmonella*; detection;

XX KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;

XX KW rotavirus; food composition; pharmaceutical composition; gene; ds.

XX OS Bifidobacterium longum.

XX OS Synthetic.

XX PN EP1227152-A1.

XX PD 31-JUL-2002.

XX PF 30-JAN-2001; 2001EP-00102050.

XX PR 30-JAN-2001; 2001EP-00102050.

XX PA (NEST) SOC PROD NESTLE SA.

XX DR WPI; 2002-668397/72.

XX PT Novel polynucleotide comprising Bifidobacterium genome sequence useful as

XX PT a probe or primer for detecting and/or identifying Bifidobacterium longum

XX PT in a biological sample.

XX PS Disclosure; SEQ ID NO 1100; 80pp; English.

XX CC The present invention describes a polynucleotide (I) comprising a

XX CC sequence of a Bifidobacterium genome selected from the nucleotide

XX CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at

XX CC least 90% identity or which hybridises with the sequences given in

XX CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a

XX CC fusion protein, comprising a sequence selected from 1097 sequences given

XX in ABP65258 to ABP6354 ligated in frame to a polynucleotide encoding a

XX CC heterologous polypeptide. (I) has antidiarrheic and antibacterial

XX CC activities, and can be used as an inhibitor of *Salmonella*. (I) which is

XX CC a probe) is useful for the detection and/or identification of

XX CC Bifidobacterium longum in a biological sample. A carrier containing the

XX CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618) can be

XX CC used for preventing and/or treating diarrhoea brought about by pathogenic

XX CC bacteria and/or rotavirus. The carrier is a food composition selected

XX CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented

XX CC products, ice-creams, fermented cereal based products, milk based

XX CC powders, infant formula, pet food or a pharmaceutical composition

XX CC selected from tablets, liquid bacterial suspensions, dried oral

XX CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.

XX CC (I) is useful in DNA arrays or chips to carry out analysis of the

XX CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent

XX CC Bifidobacterium related nucleotide sequences given in the Sequence

XX CC listing from the present invention but not mentioned further within the

XX CC specification. N.B. The sequence data for this patent is not represented

XX CC in the printed specification but is based on sequence information

XX CC supplied by the European Patent Office

XX SQ Sequence 349980 BP; 69975 A; 105045 C; 104394 G; 70566 T; 0 U; 0 Other;

Query Match 66.2%; Score 17.2; DB 6; Length 349980;

Best Local Similarity 86.4%; Pred. No. 3.6e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGCGGAGGCGCGCAATCTACGAA 26

DB 216363 CGCGGAGGCGCGTATCGAGAA 216384

RESULT 28

ABQ14535

ID ABQ14535 standard; DNA; 530 BP.

XX AC ABQ14535;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 1126.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX KW drug; side effect; cancer; central nervous system; cardiovascular;

XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;

XX KW SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-EP010074.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PR 05-SEP-2000; 2000DE-01044543.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX DR WPI; 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful for

XX PT diagnosis and prognosis, comprises selective hybridization of amplicons

XX PT from chemically treated DNA.

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of

XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

XX CC genomic sample of DNA. The sample is treated chemically to convert